



08-17-00

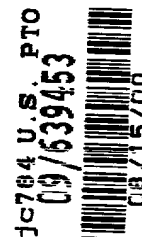
A

PATENT

Attorney's Docket Number: 7705.0002-03

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

ASSISTANT COMMISSIONER FOR PATENTS
Washington, D.C. 20231



CERTIFICATE OF EXPRESS MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service's "Express Mail Post Office to Addressee" service under 37 CFR § 1.10, in an envelope addressed to: BOX PATENT APPLICATION, Assistant Commissioner For Patents, Washington, D.C. 20231, on August 15, 2000.

Express Mail Label No. **EL 589694110 US**

Signed:

Linda Phillips
Linda Phillips

Prior Application:

Art Unit: 1636

Examiner: W. Sandals

SIR: This is a request for filing a Divisional Application under 37 C.F.R. § 1.53(b) of pending prior application Serial No. 08/942,806, filed October 2, 1997, of Arthur T. Sands, Glenn A. Friedrich, Brian Zambrowicz, and Allan Bradley, for AN INDEXED LIBRARY OF CELLS CONTAINING GENOMIC MODIFICATIONS AND METHODS OF MAKING AND UTILIZING THE SAME.

1. ☒ Enclosed is a complete copy of the prior application including the oath or Declaration and drawings, if any, as originally filed. On information and belief, I hereby verify that the attached papers are true copies of prior application Serial No. 08/942,806 as originally filed on October 2, 1997, and of the Declaration filed March 3, 1998.
2. ☐ Enclosed is a substitute specification under 37 C.F.R. § 1.125.
3. ☒ Cancel Claims 7 and 9 to 28.
4. ☐ A Preliminary Amendment is enclosed.
5. ☒ The filing fee is calculated on the basis of the claims existing in the prior application as amended at 3 and 4 above.

LAW OFFICES

FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.
STANFORD RESEARCH PARK
700 HANSEN WAY
PALO ALTO, CALIF. 94304
650-849-6600

Basic Application Filing Fee					\$690	\$ 690.00
	Number of Claims		Basic	Extra Claims		
Total Claims	7	-	20	0	x \$18	
Independent Claims	1	-	3	0	x \$78	
[X] Presentation of Multiple Dep. Claim(s)					+\$260	260.00
					Subtotal	\$ 950.00
					Reduction by 1/2 if small entity	-
					TOTAL APPLICATION FILING FEE	\$ 950.00

6. ■ Please do not charge the filing fee to Deposit Account No. 06-0916.
7. ■ The Commissioner is hereby authorized to charge any fees which may be required including fees due under 37 C.F.R. § 1.16 and any other fees due under 37 C.F.R. § 1.17, or credit any overpayment during the pendency of this application to Deposit Account No. 06-0916.
8. ■ Amend the specification at page 1, line 3, after "The present application" by inserting:

--is a divisional of U.S. Application Serial No. 08/942,806, filed October 2, 1997, which--.
9. ■ Since applicants intend the present divisional application to have the same disclosure as parent application Serial No. 08/942,806, that application has been incorporated by reference into this application in an abundance of caution in the event that any of the disclosure of Serial No. 08/942,806 is inadvertently omitted in this submission. That incorporation by reference should not necessitate a new oath or declaration, since the declaration (a copy of which is enclosed) already was executed for the disclosure of Serial No. 08/942,806.
10. □ New formal drawings are enclosed.
11. ■ The prior application is assigned of record to: Lexicon Genetics Incorporated, 4000 Research Forest Drive, The Woodlands, TX 77381.

LAW OFFICES

FINNEGAN, HENDERSON,
 FARABOW, GARRETT
 & DUNNER, L.L.P.
 STANFORD RESEARCH PARK
 700 HANSEN WAY
 PALO ALTO, CALIF. 94304
 650-849-6600

12. ☐ Priority of application Serial No. _____, filed on _____ in _____
 _____ (country) is claimed under 35 U.S.C. § 119. A certified copy
☐ is enclosed or ☐ is on file in the prior application.
13. ☐ A verified statement claiming small entity status
☐ is enclosed or ☐ is on file in the prior application.
14. ☒ The power of attorney in the prior application is to at least one of the
 following: FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER,
 L.L.P., Douglas B. Henderson, Reg. No. 20,291; Ford F. Farabow, Jr., Reg.
 No. 20,630; Arthur S. Garrett, Reg. No. 20,338; Donald R. Dunner, Reg.
 No. 19,073; Brian G. Brunsvold, Reg. No. 22,593; Tipton D. Jennings, IV,
 Reg. No. 20,645; Jerry D. Voight, Reg. No. 23,020; Laurence R. Hefter, Reg.
 No. 20,827; Kenneth E. Payne, Reg. No. 23,098; Herbert H. Mintz, Reg.
 No. 26,691; C. Larry O'Rourke, Reg. No. 26,014; Albert J. Santorelli, Reg.
 No. 22,610; Michael C. Elmer, Reg. No. 25,857; Richard H. Smith, Reg.
 No. 20,609; Stephen L. Peterson, Reg. No. 26,325; John M. Romary, Reg.
 No. 26,331; Bruce C. Zotter, Reg. No. 27,680; Dennis P. O'Reilley, Reg.
 No. 27,932; Allen M. Sokal, Reg. No. 26,695; Robert D. Bajefsky, Reg.
 No. 25,387; Richard L. Stroup, Reg. No. 28,478; David W. Hill, Reg.
 No. 28,220; Thomas L. Irving, Reg. No. 28,619; Charles E. Lipsey, Reg.
 No. 28,165; Thomas W. Winland, Reg. No. 27,605; Basil J. Lewris, Reg.
 No. 28,818; Martin I. Fuchs, Reg. No. 28,508; E. Robert Yoches, Reg.
 No. 30,120; Barry W. Graham, Reg. No. 29,924; Susan Haberman Griffen,
 Reg. No. 30,907; Richard B. Racine, Reg. No. 30,415; Thomas H. Jenkins,
 Reg. No. 30,857; Robert E. Converse, Jr., Reg. No. 27,432; Clair X. Mullen,
 Jr., Reg. No. 20,348; Christopher P. Foley, Reg. No. 31,354; John C. Paul,
 Reg. No. 30,413; David M. Kelly, Reg. No. 30,953; Kenneth J. Meyers, Reg.
 No. 25,146; Carol P. Einaudi, Reg. No. 32,220; Walter Y. Boyd, Jr., Reg.
 No. 31,738; Steven M. Anzalone, Reg. No. 32,095; Jean B. Fordis, Reg.
 No. 32,984; Roger D. Taylor, Reg. 28,992; Barbara C. McCurdy, Reg.
 No. 32,120; James K. Hammond, Reg. No. 31,964; Richard V. Burgujian,
 Reg. No. 31,744; J. Michael Jakes, Reg. No. 32,824; Thomas W. Banks,
 Reg. No. 32,719; Christopher P. Isaac, Reg. No. 32,616; Bryan C. Diner,
 Reg. No. 32,409; M. Paul Barker, Reg. No. 32,013; Andrew Chanhon Sonu,
 Reg. No. 33,457; David S. Forman, Reg. No. 33,694; Vincent P. Kovalick,
 Reg. No. 32,867; James W. Edmondson, Reg. No. 33,871; Michael R.
 McGurk, Reg. No. 32,045; Joann M. Neth, Reg. No. 36,363; Gerson S.
 Panitch, Reg. No. 33,751; Cheri M. Taylor, Reg. No. 33,216; Charles E. Van

LAW OFFICES

FINNEGAN, HENDERSON,
 FARABOW, GARRETT
 & DUNNER, L.L.P.
 STANFORD RESEARCH PARK
 700 HANSEN WAY
 PALO ALTO, CALIF. 94304
 650-849-6600


Horn, Reg. No. 40,266; Linda A. Wadler, Reg. No. 33,218; Jeffrey A. Berkowitz, Reg. No. 36,743; Michael R. Kelly, Reg. No. 33, 921; and James B. Monroe, Reg. No. 33,971; Doris Johnson Hines, Reg. No. 34,629; Allen R. Jensen, Reg. No. 28,224; Lori Ann Johnson, Reg. No. 34,498; and David A. Manspeizer, Reg. No. 37,540.

15. ☐ The power appears in the original declaration of the prior application.
16. ☒ Since the power does not appear in the original declaration, a copy of the Revocation of Power of Attorney And Grant of New Power of Attorney in the prior application Serial No. 08/942,806 is enclosed.
17. ☒ Please address all correspondence to FINNEGAN, HENDERSON, FARABOW, GARRETT and DUNNER, L.L.P., 1300 I Street, N.W., Washington, D.C. 20005-3315.
18. ☐ Recognize as associate attorney _____

 (name, address & Reg. No.)
19. ☐ Also enclosed is _____

PETITION FOR EXTENSION. If any extension of time is necessary for the filing of this application, including any extension in the parent application, serial no. 08/942,806, filed October 2, 1997, for the purpose of maintaining copendency between the parent application and this application, and such extension has not otherwise been requested, such an extension is hereby requested, and the Commissioner is authorized to charge necessary fees for such an extension to our Deposit Account No. 06-0916. A duplicate copy of this paper is enclosed for use in charging the deposit account.

FINNEGAN, HENDERSON, FARABOW,
GARRETT & DUNNER, L.L.P.

By: 
 M. Paul Barker
 Reg. No.: 32,013

Date: August 15, 2000

**AN INDEXED LIBRARY OF CELLS CONTAINING GENOMIC MODIFICATIONS
AND METHODS OF MAKING AND UTILIZING THE SAME**

The present application is a continuation-in-part of
5 U.S. Applications Ser. Nos. 08/726,867, filed October 4,
1996, and 08/728,963, filed October 11, 1996. The
application also claims priority to U.S. Application Ser. No.
08/907,598, filed August 8, 1997. The disclosures of the
above applications are herein incorporated by reference.

10

1.0. FIELD OF THE INVENTION

The invention relates to an indexed library of
genetically altered cells and methods of organizing the cells
into an easily manipulated and characterized Library. The
15 invention also relates to methods of making the library,
vectors for making insertion mutations in genes, methods of
gathering sequence information from each member clone of the
Library, and methods of isolating a particular clone of
interest from the Library.

20

2.0. BACKGROUND OF THE INVENTION

The general technologies of targeting mutations into the
genome of cells, and the process of generating mouse lines
from genetically altered embryonic stem (ES) cells with
specific genetic lesions are well known (Bradley, 1991, Cur.
25 Opin. Biotech. 2:823-829). A random method of generating
genetic lesions in cells (called gene, or promoter, trapping)
has been developed in parallel with the targeted methods of
genetic mutation (Allen et al., 1988 Nature 333(6176):852-
30 855; Brenner et al., 1989, Proc. Natl. Acad. Sci. U.S.A.
86(14):5517-5521; Chang et al., 1993, Virology 193(2):737-
747; Friedrich and Soriano, 1993, Insertional mutagenesis by
retroviruses and promoter traps in embryonic stem cells, p.
681-701. In Methods Enzymol., vol. 225., P. M. Wassarman and
35 M. L. DePamphilis (ed.), Academic Press, Inc., San Diego;
Friedrich and Soriano, 1991, Genes Dev. 5(9):1513-1523;
Gossler et al., 1989, Science 244(4903):463-465; Kerr et al.,

1989, Cold Spring Harb. Symp. Quant. Biol. 2:767-776; Reddy et al., 1991, J Virol. 65(3):1507-1515; Reddy et al., 1992, Proc. Natl. Acad. Sci. U.S.A. 89(15):6721-6725; Skarnes et al., 1992, Genes Dev. 6(6):903-918; von Melchner and Ruley, 5 1989, J. Virol. 63(8):3227-3233; Yoshida et al., 1995, Transgen. Res. 4:277-287). Gene trapping provides a means to create a collection of random mutations by inserting fragments of DNA into transcribed genes. Insertions into transcribed genes are selected over the background of total 10 insertions since the mutagenic DNA encodes an antibiotic resistance gene or some other selectable marker. The selectable marker lacks its own promoter and enhancer and must be expressed by the endogenous sequences that flank the marker after it has integrated. Using this approach, 15 transcription of the selectable marker is activated and the cell gene is concurrently mutated. This type of strict selection makes it possible to easily isolate thousands of ES cell colonies, each with a unique mutagenic insertion.

Collecting mutants on a large-scale has been a powerful 20 genetic technique commonly used for organisms which are more amenable to such analysis than mammals. These organisms, such as *Drosophila melanogaster*, yeast *Saccharomyces cerevisiae*, and plants such as *Arabidopsis thaliana* are small, have short generation times and small genomes (Bellen et al., 25 1989, Genes Dev. 3(9):1288-1300; Bier et al., 1989, Genes Dev. 3(9):1273-1287; Hope, 1991, Develop. 113(2):399-408. These features allow an investigator to rear many thousands or millions of different mutant strains without requiring unmanageable resources. However, these type of organisms 30 have only limited value in the study of biology relevant to human physiology and health. It is therefore important to have the power of large-scale genetic analysis available for the study of a mammalian species that can aid in the study of human disease. Given that the entire human genome is 35 presently being sequenced, the comprehensive genetic analysis of a related mammalian species will provide a means to determine the function of genes cloned from the human genome.

At present, rodents, and particularly mice, provide the best model for genetic manipulation and analysis of mammalian physiology.

Gene trapping has been used as an analytical tool to identify genes and regulatory regions in a variety of animal cell types. One system that has proved particularly useful is based on the use of ROSA (reverse orientation splice acceptor) retroviral vectors (Friedrich and Soriano, 1991 and 1993).

10 The ROSA system can generate mutations that result in a detectable homozygous phenotype with a high frequency. About 50% of all the insertions caused embryonic lethality. The specifically mutated genes may easily be cloned since the gene trapping event produces a fusion transcript. This
15 fusion transcript has trapped exon sequences appended to the sequences of the selectable marker allowing the latter to be used as a tag in polymerase chain reaction (PCR)-based protocols, or by simple cDNA cloning. Examples of genes isolated by these methods include a transcription factor
20 related to human TEF-1 (transcription enhancer factor-1) which is required in the development of the heart (Chen et al., 1994, Genes Devel. 8:2293-2301. Another (spock), is distantly related to yeast genes encoding secretion proteins and is important during gastrulation.

25 The above experiments have established that the ROSA system is an effective analytical tool for genetic analysis in mammals. However, the structure of many ROSA vectors selects for the "trapping" of 5' exons which, in many cases, do not encode proteins. Such a result is adequate where one
30 wishes to identify and eventually clone control (i.e., promoter or enhancer) sequences, but is not optimal where the generation of insertion-inactivated null mutations is desired, and relevant coding sequence is needed. Thus, the construction of large-scale mutant (preferably null mutant)
35 libraries requires the use of vectors that have been designed to select for insertion events that have occurred within the coding region of the mutated genes as well as vectors that

are not limited to detecting insertions into expressed genes.

3.0. SUMMARY OF THE INVENTION

An object of the present invention is to provide a set
5 of genetically altered cells (the 'Library'). The genetic
alterations are of sufficient randomness and frequency such
that the combined population of cells in the Library
represent mutations in essentially every gene found in the
cell's genome. The Library is used as a source for obtaining
10 specifically mutated cells, cell lines derived from the
individually mutated cells, and cells for use in the
production of transgenic non-human animals.

A further object is to provide the vectors, both DNA and
retroviral based, that may be used to generate the Library.
15 Typically, at least two distinct vector designs will be used
in order to mutate genes that are actively expressed in the
target cell, and genes that are not expressed in the target
cell. Combining the mutant cells obtained using both types
of vectors best ensures that the Library provides a
20 comprehensive set of gene mutations.

A particularly useful vector class contemplated by the
present invention includes a vector for inserting foreign
exons into animal cell transcripts that comprises a
selectable marker, a promoter element operatively positioned
25 5' to the selectable marker, a splice donor site operatively
positioned 3' to the selectable marker, and a second
mutagenic foreign polynucleotide sequence located upstream
from the promoter element that disrupts, or otherwise
"poisons", the splicing or read-through expression of the
30 endogenous cellular transcript. Typically, the mutagenic
foreign polynucleotide sequence may incorporate a
polyadenylation (pA) site, a nested set of stop codons in
each of the three reading frames, splice acceptor and splice
donor sequences in operable combination, a mutagenic exon, or
35 any mixture of mutagenic features that effectively prevent
the expression of the cellular gene. For example, a
polyadenylation sequence may be incorporated in addition to

or in lieu of the splice donor sequence. A preferred organization for the mutagenic polynucleotide sequence comprises a polyadenylation site positioned upstream from a selectable marker which is in turn located upstream from a splice acceptor sequence. Preferably, such a vector does not comprise a transcription terminator or polyadenylation site operatively positioned relative to the coding region of the selectable marker, and shall not comprise a splice acceptor site operatively positioned between the promoter element and the initiation codon of said selectable marker.

An additional vector contemplated by the present invention is designed to replace the normal 3' end of an animal cell transcript with a foreign exon. Such a vector shall generally be engineered to comprise a selectable marker, a splice acceptor site operatively positioned upstream (5') from the initiation codon of the selectable marker, and a polyadenylation site operatively positioned downstream (3') from the termination codon (3' end) of the selectable marker. Preferably, the vector will not comprise a promoter element operatively positioned upstream from the coding region of the selectable marker, and will not comprise a splice donor sequence operatively positioned between the 3' end of the coding region of the selectable marker and the polyadenylation site.

Yet another vector contemplated by the present invention is a vector designed to insert a mutagenic foreign polynucleotide sequence within an animal cell transcript (i.e., the foreign polynucleotide sequence is flanked on both sides by endogenous exons). As described above, the mutagenic foreign polynucleotide sequence may be any sequence that disrupts the normal expression of the gene into which the vector has integrated. Optionally, the vector may additionally incorporate a selectable marker, a splice acceptor site operatively positioned 5' to the initiation codon of the selectable marker, a splice donor site operatively positioned 3' to said selectable marker. Preferably, this vector shall not comprise a polyadenylation

site operatively positioned 3' to the coding region of said selectable marker, and shall not comprise a promoter element operatively positioned 5' to the coding region of said selectable marker.

5 An additional embodiment of the present invention is a library of genetically altered cells that have been treated to stably incorporate one or more types of the vectors described above. The presently described library of cultured animal cells may be made by a process comprising the
10 steps of treating (i.e., infecting, transfecting, retrotransposing, or virtually any other method of introducing polynucleotides into a cell) a population of cells to stably integrate a vector that mediates the splicing of a foreign exon internal to a cellular transcript,
15 transfecting another population of cells to stably integrate a vector that mediates the splicing of a foreign exon 5' to an exon of a cellular transcript, and selecting for transduced cells that express the products encoded by the foreign exons.

20 Alternatively, an additional embodiment of the present invention describes a mammalian cell library made by a method comprising the steps of: transfecting a population of cells with a vector capable of expressing a selectable marker in the cell only after the vector inserts into the host genome;
25 transfecting or infecting a population of cells with a vector containing a selectable marker that is substantially only expressed by cellular control sequences (after the vector integrates into the host cells genome); and growing the transfected cells under conditions that select for the
30 expression of the selectable marker.

In an additional embodiment of the present invention, the two populations of transfected cells will be individually grown under selective conditions, and the resulting mutated population of cells collectively comprises a substantially
35 comprehensive library of mutated cells.

In an additional embodiment of the present invention, the individual mutant cells in the library are separated and

clonally expanded. Additionally, the clonally expanded mutant cells may then be analyzed to ascertain the DNA sequence, or partial DNA sequence of the mutated host gene.

The presently described methods of making, organizing, and indexing libraries of mutated animal cells are also broadly applicable to virtually any eukaryotic cells that may be genetically manipulated and grown in culture.

The invention provides for sequencing every gene mutated in the Library. The resulting sequence database subsequently serves as an index for the library. In essence, every cell line in the Library is individually catalogued using the partial sequence information. The resulting sequence is specific for the mutated gene since the present methods are designed to obtain sequence information from exons that have been spliced to the marker sequence. Since the coverage of the mutagenesis is preferably the entire set of genes in the genome, the resulting Library sequence database contains sequence from essentially every gene in the cell. From this database, a gene of interest can be identified. Once identified, the corresponding mutant cell may be withdrawn from the Library based on cross reference to the sequence data.

An additional embodiment of the invention provides for methods of isolating mutations of interest from the Library. Two methods are proposed for obtaining individual mutant cell lines from the Library. The first provides a scheme where clones of the cells generated using the above vectors are pooled into sets of defined size. Using the procedure described below which utilizes reverse transcription (RT) and polymerase chain reaction (PCR), a cell line with a mutation in a gene whose sequence is partly or wholly known is isolated from organized sets of these pools. A few rounds of this screening procedure results in the isolation of the desired individual cell line.

A second procedure involves the sequencing of regions flanking the vector insertion sites in the various cells in the library. The sequence database generated from these data

effectively constitutes an index of the clones in the library that may be used to identify cells having mutations in specific genes.

5 **4.0. DESCRIPTION OF THE FIGURES**

Figure 1. Shows a diagrammatic representation of 5 different vectors that are generally representative of the type of vectors that may be used in the present invention.

10 Figure 2. Shows a general strategy for identifying "trapped" cellular sequences by PCR analysis of the cellular exons that flank the foreign intron introduced by the VICTR 2 vector.

Figure 3 shows a PCR based strategy for identifying tagged
15 genes by chromosomal location.

Figure 4. Is a diagrammatic representation of a strategy of identifying or indexing the specific clones in the library via PCR analysis and sequencing of mRNA samples obtained from
20 the cells in the library.

Figure 5. Is a diagrammatic representation of a method of isolating positive clones by screening pooled mutant cell clones.

25

Figure 6. Partial nucleic acid or predicted amino acid sequence data from 9 clones (OST1-9) isolated using the described techniques aligned with similar sequences from previously characterized genes.

30

Figure 7. Provides a diagrammatic representation of VICTRs 3 and 20 as well as the transcripts that result after integration into a hypothetical region of the target cell genome (i.e., "Wildtype Locus).

35

Figure 8. Provides a representative list of a portion of the known genes that have been identified using the disclosed

methods and technology.

5.0. DETAILED DESCRIPTION OF THE INVENTION

The present invention describes a novel indexed library
5 containing a substantially comprehensive set of mutations in
the host cell genome, and methods of making and using the
same. The presently described Library comprises as a set of
cell clones that each possess at least one mutation (and
preferably a single mutation) caused by the insertion of DNA
10 that is foreign to the cell. For the purposes of the present
invention, "foreign" polynucleotide sequences can be any
sequences that are newly introduced to a cell, do not
naturally occur in the cell at the engineered region of the
chromosome, or occur in the cell but are not organized to
15 provide an identical function to that provided in the
engineered vector.

The particularly novel features of the Library include
the methods of construction, and indexing. To index the
library, the mutant cells of the library are clonally
20 expanded and each mutated gene is at least partially
sequenced. The Library thus provides a novel tool for
assessing the specific function of a given gene. The
insertions cause a mutation which allow for essentially every
gene represented in the Library to be studied using genetic
25 techniques either *in vitro* or *in vivo* (via the generation of
transgenic animals). For the purposes of the present
invention, the term "essentially every gene" shall refer to
the statistical situation where there is generally at least
about a 70 percent probability that the genomes of cells used
30 to construct the library collectively contain at least one
inserted vector sequence in each gene, preferably a 85
percent probability, and more specifically at least about a
95 percent probability as determined by a standard Poisson
distribution.

35 Also for the purposes of the present invention the term
"gene" shall refer to any and all discrete coding regions of
the cell's genome, as well as associated noncoding and

regulatory regions. Additionally, the term operatively positioned shall refer to the control elements or genes that are provided with the proper orientation and spacing to provide the desired or indicated functions of the control
5 elements or genes.

For the purposes of the present invention, a gene is "expressed" when a control element in the cell mediates the production of functional or detectable levels of mRNA encoded by the gene, or a selectable marker inserted therein. A gene
10 is not expressed where the control element in the cell is absent, has been inactivated, or does not mediate the production of functional or detectable levels of mRNA encoded by the gene, or a selectable marker inserted therein.

15 5.1. Vectors used to build the Library

A number of investigators have developed gene trapping vectors and procedures for use in mouse and other cells (Allen et al., 1988; Bellen et al., 1989, Genes Dev. 3(9):1288-1300; Bier et al., 1989, Genes Dev. 3(9):1273-1287;
20 Bonnerot et al., 1992, J Virol. 66(8):4982-4991; Brenner et al., 1989; Chang et al., 1993; Friedrich and Soriano, 1993; Friedrich and Soriano, 1991; Goff, 1987, Methods Enzymol. 152:469-481; Gossler et al.; Hope, 1991; Kerr et al., 1989; Reddy et al., 1991; Reddy et al., 1992; Skarnes et al., 1992;
25 von Melchner and Ruley; Yoshida et al., 1995). The gene trapping system described in the present invention is based on significant improvements to the published SA (splice acceptor) DNA vectors and the ROSA (reverse orientation, splice acceptor) retroviral vectors (Chen et al., 1994;
30 Friedrich and Soriano, 1991 and 1993). The presently described vectors also use a selectable marker called β geo. This gene encodes a protein which is a fusion between the β -galactosidase and neomycin phosphotransferase proteins. The presently described vectors place a splice acceptor sequence
35 upstream from the β geo gene and a poly-adenylation signal sequence downstream from the marker. The marker is integrated after transfection by, for example,

electroporation (DNA vectors), or retroviral infection, and gene trap events are selected based on resistance to G418 resulting from activation of β geo expression by splicing from the endogenous gene into the ROSA splice acceptor. This type
5 of integration disrupts the transcription unit and preferably results in a null mutation at the locus.

Although gene trapping has proven a useful analytical tool, the present invention contemplates gene trapping on a large scale. The vectors utilized in the present invention
10 have been engineered to overcome the shortcomings of the early gene trap vector designs, and to facilitate procedures allowing high throughput. In addition, procedures are described that allow the rapid and facile acquisition of sequence information from each trapped cDNA which may be
15 adapted to allow complete automation. These latter procedures are also designed for flexibility so that additional molecular information can easily be obtained subsequently. The present invention therefore incorporates gene trapping into a larger and unique tool. A specially
20 organized set of gene trap clones that provide a novel and powerful new tool of genetic analysis.

The presently described vectors are superficially similar to the ROSA family of vectors, but constitute significant improvements and provide for additional features
25 that are useful in the construction and indexing of the Library. Typically, gene trapping vectors are designed to detect insertions into transcribed gene regions within the genome. They generally consist of a selectable marker whose normal expression is handicapped by exclusion of some element
30 required for proper transcription. When the vector integrates into the genome, and acquires the necessary element by juxtaposition, expression of the selectable marker is activated. When such activation occurs, the cell can survive when grown in the appropriate selective medium which
35 allows for the subsequent isolation and characterization of the trapped gene. Integration of the gene trap generally causes the gene at the site of integration to be mutated.

Some gene trapping vectors have a splice acceptor preceding a selectable marker and a poly-adenylation signal following the selectable marker, and the selectable marker gene has its own initiator ATG codon. Using this arrangement, the fusion transcripts produced after integration generally only comprise exons 5' to the insertion site to the known marker sequences. Where the vector has inserted into the 5' region of the gene, it is often the case that the only exon 5' to the vector is a non-coding exon. Accordingly, the sequences obtained from such fusions do not provide the desired sequence information about the relevant gene products. This is because untranslated sequences are generally less well conserved than coding sequences.

To compensate for the short-comings of earlier vectors, the vectors of the present invention have been designed so that 3' exons are appended to the fusion transcript by replacing the poly-adenylation and transcription termination signals of earlier ROSA vectors with a splice donor (SD) sequence. Consequently transcription and splicing generally results in a fusion between all or most of the endogenous transcript and the selectable marker exon, for example β geo, neomycin (neo) or puromycin (puro). The exon sequences immediately 3' to the selectable marker exon may then be sequenced and used to establish a database of expressed sequence tags. The presently described procedures will typically provide approximately 200 nucleotides of sequence, or more. These sequences will generally be coding and therefore informative. The prediction that the sequence obtained will be from coding region is based on two factors. First, gene trap vectors are generally found near the 5' end of the gene immediately after untranslated exons because the method selects for integration events that place the initiator ATG of the selectable marker as the first encountered, and thus used, for translation. Second, mammalian transcripts have short 5' untranslated regions (UTRs) which are typically between 50 and 150 nucleotides in length.

The obtained sequence information also provides a ready source of probes that may be used to isolate the full-length gene or cDNA from the host cell, or as heterologous probes for the isolation of homologous genes in other species.

5 Internal exons in mammalian transcripts are generally quite small, on the average 137 bases with few over 300 bases. Consequently, a large internal exon may be spliced less efficiently. Thus, the presently described vectors have been designed to sandwich relatively small selectable markers
10 (for example: neo, ~800 bases, or a smaller drug resistance gene such as puro, ~600 bases) between the requisite splicing elements to produce relatively small exons. Exons of this size are more typical of mammalian exons and do not present undue problems for the splicing machinery of the cell. Such
15 a design consideration is novel to the presently disclosed gene trapping vectors. Accordingly, an additional embodiment of the claimed vectors is that the respective splice acceptor and splice donor sites are engineered such that they are operatively positioned close to the ends of the selectable
20 marker coding region (the region spanning from the initiation codon to the termination codon). Generally, the splice acceptor or splice donor sequences shall appear within about 80 bases from the nearest end of the selectable marker coding region, preferably within about 50 bases from the nearest end
25 of the coding region, more preferably within about 30 bases from the nearest end of the coding regions and specifically within about 20 bases of the nearest end of the selectable marker coding region.

The new vectors are represented in retroviral form in
30 Figure 1. They are used by infecting target cells with retroviral particles such that the proviruses shown in the schematic can be found in the genome of the target. These vectors are called VICTR which is an acronym for "viral constructs for trapping".

35 The presently described retroviral vectors may be used in conjunction with retroviral packaging cell lines such as those described in U.S. Patent No. 5,449,614 ("'614 patent")

issued September 12, 1995, herein incorporated by reference. Where non-mouse animal cells are to be used as targets for generating the described libraries, packaging cells producing retrovirus with amphotropic envelopes will generally be employed to allow infection of the host cells.

The mutagenic gene trap DNA may also be introduced into the target cell genome by various transfection techniques which are familiar to those skilled in the art such as electroporation, lipofection, calcium phosphate precipitation, infection, retrotransposition, and the like. Examples of such techniques may be found in Sambrook et al. (1989) Molecular Cloning Vols. I-III, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, and Current Protocols in Molecular Biology (1989) John Wiley & Sons, all Vols. and periodic updates thereof, herein incorporated by reference. The transfected versions of the retroviral vectors are typically plasmid DNA molecules containing DNA cassettes comprising the described features between the retroviral LTRs.

The vectors VICTR 1 and 2 (Fig. 1) are designed to trap genes that are transcribed in the target cell. To trap genes that are not expressed in the target cell, gene trap vectors such as VICTR 3, 4 and 5 (described below) are provided. These vectors have been engineered to contain a promoter element capable of initiating transcription in virtually any cell type which is used to transcribe the coding sequence of the selectable marker. However, in order to get proper translation of the marker product, and thus render the cell resistant to the selective antibiotic, a polyadenylation signal and a transcription termination sequence must be provided. Vectors VICTR 3 through 5 are constructed such that an effective polyadenylation signal can only be provided by splicing with an externally provided downstream exon that contains a poly-adenylation site. Therefore, since the selectable marker coding region ends only in a splice donor sequence, these vectors must be integrated into a gene in order to be properly expressed. In essence, these vectors

append the foreign exon encoding the marker to the 5' end of an endogenous transcript. These events will tag genes and create mutations that are used to make clones that will become part of the Library.

5 With the above design considerations, the VICTR series of vectors, or similarly designed and constructed vectors, have the following features. VICTR 1 is a terminal exon gene trap. VICTR 1 does not contain a control region that effectively mediates the expression of the selectable marker
10 gene. Instead, the coding region of the selectable marker contained in VICTR 1, in this case encoding puromycin resistance (but which can be any selectable marker functional in the target cell type), is preceded by a splice acceptor sequence and followed by a polyadenylation addition signal
15 sequence. The coding region of the puro gene has an initiator ATG which is downstream and adjacent to a region of sequence that is most favorable for translation initiation in eukaryotic cells - the so called Kozak consensus sequence (Kozak, 1989, J. Cell, Biol. 108(2):229-241). With a Kozak
20 sequence and an initiator ATG, the puro gene in VICTR 1 is activated by integrating into the intron of an active gene, and the resulting fusion transcript is translated beginning at the puromycin initiation (ATG/AUG) codon. However, terminal gene trap vectors need not incorporate an initiator
25 ATG codon. In such cases, the gene trap event requires splicing and the translation of a fusion protein that is functional for the selectable marker activity. The inserted puromycin coding sequence must therefore be translated in the same frame as the "trapped" gene.

30 The splice acceptor sequence used in VICTR 1 and other members of the VICTR series is derived from the adenovirus major late transcript splice site located at the intron 1/exon 2 boundary. This sequence contains a polypyrimidine stretch preceding the AG dinucleotide which denotes the
35 actual splice site. The presently described vectors contemplate the use of any similarly derived splice acceptor sequence. Preferably, the splice acceptor site will only

rarely, if ever, be involved in alternative splicing events.

The polyadenylation signal at the end of the *puro* gene is derived from the bovine growth hormone gene. Any similarly derived polyadenylation signal sequence could be used if it contains the canonical AATAAA and can be demonstrated to terminate transcription and cause a polyadenylate tail to be added to the engineered coding exons.

VICTR 2 is a modification of VICTR 1 in which the polyadenylation signal sequence is removed and replaced by a splice donor sequence. Like VICTR 1, VICTR 2 does not contain a control region that effectively mediates the expression of the selectable marker gene. Typically, the splice donor sequence to be employed in a VICTR series vector shall be determined by reference to established literature or by experimentation to identify which sequences properly initiate splicing at the 5' end of introns in the desired target cell. The specifically exemplified sequence, AGGTAAGT, results in splicing occurring in between the two G bases. Genes trapped by VICTR 2 splice upstream exons onto the *puro* exon and downstream exons onto the end of the *puro* exon. Accordingly, VICTR 2 effectively mutates gene expression by inserting a foreign exon in-between two naturally occurring exons in a given transcript. Again, the *puro* gene may or may not contain a consensus Kozak translation initiation sequence and properly positioned ATG initiation codon. As discussed above, gene trapping by VICTR 1 and VICTR 2 requires that the mutated gene is expressed in the target cell line. By incorporating a splice donor into the VICTR traps, transcript sequences downstream from the gene trap insertion can be determined. As described above, these sequences are generally more informative about the gene mutated since they are more likely to be coding sequences. This sequence information is gathered according to the procedures described below.

VICTR 3, VICTR 4 and VICTR 5 are gene trap vectors that do not require the cellular expression of the endogenous

trapped gene. The VICTR vectors 3 through 5 all comprise a promoter element that ensures that transcription of the selectable marker would be found in all cells that have taken up the gene trap DNA. This transcription initiates from a promoter, in this case the promoter element from the mouse phosphoglycerate kinase (PGK) gene. However, since the constructs lack a polyadenylation signal there can be no proper processing of the transcript and therefore no translation. The only means to translate the selectable marker and get a resistant cell clone is by acquiring a polyadenylation signal. Since polyadenylation is known to be concomitant with splicing, a splice donor is provided at the end of the selectable marker. Therefore, the only positive gene trap events using VICTR 3 through 5 will be those that integrate into a gene's intron such that the marker exon is spliced to downstream exons that are properly polyadenylated. Thus genes mutated with the VICTR vectors 3 through 5 need not be expressed in the target cell, and these gene trap vectors can mutate all genes having at least one intron. The design of VICTR vectors 3 through 5 requires a promoter element that will be active in the target cell type, a selectable marker and a splice donor sequence. Although a specific promoter was used in the specific embodiments, it should be understood that appropriate promoters may be selected that are known to be active in a given cell type. Typically, the considerations for selecting the splice donor sequence are identical to those discussed for VICTR 2, *supra*.

VICTR 4 differs from VICTR 3 only by the addition of a small exon upstream from the promoter element of VICTR 4. This exon is intended to stop normal splicing of the mutated gene. It is possible that insertion of VICTR 3 into an intron might not be mutagenic if the gene can still splice between exons, bypassing the gene trap insertion. The exon in VICTR 4 is constructed from the adenovirus splice acceptor described above and the synthetic splice donor also described above. Stop codons are placed in all three reading frames in the exon, which is about 100 bases long. The stops would

truncate the endogenous protein and presumably cause a mutation.

A conceptually similar alternative design uses a terminal exon like that engineered into VICTR 5. Instead of a splice donor, a polyadenylation site is used to terminate transcription and produce a truncated message. Stops in all three frames are also provided to truncate the endogenous protein as well as the resulting transcript.

VICTR 20 is a modified version of VICTR 3 that incorporates a polyadenylation site 5' to the PGK promoter, the IRES β geo sequence (*i.e.*, foreign mutagenic polynucleotide sequence) 5' to the polyadenylation site, and a splice acceptor site 5' to the IRES β geo coding region. VICTR 20 additionally incorporates, in operable combination, a pair of recombinase recognition sites that flank the PGKpuroSD cassette.

All of the traps of the VICTR series are designed such that a fusion transcript is formed with the trapped gene. For all but VICTR 1, the fusion contains cellular exons that are located 3' to the gene trap insertion. All of the flanking exons may be sequenced according to the methods described in the following section. To facilitate sequencing, specific sequences are engineered onto the ends of the selectable marker (*e.g.*, puromycin coding region). Examples of such sequences include, but are not limited to unique sequences for priming PCR, and sequences complementary to the standard M13 forward sequencing primer. Additionally, stop codons are added in all three reading frames to ensure that no anomalous fusion proteins are produced. All of the unique 3' primer sequences are followed immediately by the synthetic 9 base pair splice donor sequence. This keeps the size of the exon comprising the selectable marker (*puro* gene) at a minimum to best ensure proper splicing, and positions the amplification and sequencing primers immediately adjacent to the flanking "trapped" exons to be sequenced as part of the construction of a Library database.

when any members of the VICTR series are constructed as retroviruses, the direction of transcription of the selectable marker is opposite to that of the direction of the normal transcription of the retrovirus. The reason for this organization is that the transcription elements such as the polyadenylation signal, the splice sites and the promoter elements found in the various members of the VICTR series interfere with the proper transcription of the retroviral genome in the packaging cell line. This would eliminate or significantly reduce retroviral titers. The LTRs used in the construction of the packaging cell line are self-inactivating. That is, the enhancer element is removed from the 3' U3 sequences such that the proviruses resulting from infection would not have an enhancer in either LTR. An enhancer in the provirus may otherwise affect transcription of the mutated gene or nearby genes.

Since a 'cryptic' splice donor sequence is found in the inverted LTRs, this splice donor sequence has been removed from the VICTR vectors by site specific mutagenesis. It was deemed necessary to remove this splice donor so that it would not affect the trapping splicing events.

The present disclosure also describes vectors that incorporate a new way to conduct positive selection. VICTR 3 and VICTR 20 are two examples of such vectors. Both VICTR 3 and VICTR 20, contain PGKpuroSD which must splice into exons of gene that provide a polyadenylation addition sequence in order to allow expression of the puromycin selectable marker gene. When placed in a targeting vector, PGKpuroSD allows for positive selection when targeting takes place. In addition to providing positive selection, targeted events among resistant colonies are easy to identify by the 3' RACE protocols (see section 5.2.2., *infra*) used for Omnibank production. This automated process allows for the rapid identification of targeted events. It is important that unlike SA β geo, PGKpuroSD does not require expression of the targeted gene in order to provide positive selection. In addition, VICTR 20 provides 2 potential positive selectable

markers (puro and neo). The use of two selectable markers, when a gene is expressed, provides a means to increase the targeting efficiency by requiring both selectable markers to function which is much more remote a possibility than having one selectable marker function unless there is a targeted event. The addition of a negative selection cassette to these vectors would only increase their targeting efficiency.

An additional feature that may be incorporated into the presently described vectors includes the use of recombinase recognition sequences. Bacteriophage P1 Cre recombinase and flp recombinase from yeast plasmids are two examples of site-specific DNA recombinase enzymes which cleave DNA at specific target sites (loxP sites for cre recombinase and frt sites for flp recombinase) and catalyze a ligation of this DNA to a second cleaved site. When a piece of DNA is flanked by 2 loxP or frt sites (e.g., recombinase control elements) in the same orientation, the corresponding recombinase will cause the removal of the intervening DNA sequence. When a piece of DNA is flanked by loxP or frt sites in an indirect orientation, the corresponding recombinase will essentially activate the control elements to cause the intervening DNA to be flipped into the opposite orientation. These recombinases provide powerful approaches for manipulating DNA *in situ*.

Recombinases have important applications for gene trapping and the production of a library of trapped genes. When constructs containing PGKpuroSD are used to trap genes, the fusion transcript between puromycin and sequences of the trapped gene could result in some level of protein expression from the trapped gene if translational reinitiation occurs. Another important issue is that several reports suggest that the PGK promoter can affect the expression of nearby genes. These effects may make it difficult to determine gene function after a gene trap event since one could not discern whether a given phenotype is associated with the inactivation of a gene, or the transcription of nearby genes. Both potential problems are solved by exploiting recombinase activity. When PGKpuroSD is flanked by loxP, frt, or any

other recombinase sites in the same orientation, the addition of the corresponding recombinase will result in the removal of PGKpuroSD. In this way, effects caused by PGKpuroSD fusion transcripts, or the PGK promoter, are avoided.

5 Accordingly, a vector that may be particularly useful for the practice of the present invention is VICTR 20. This vector replaces the terminal exon of VICTR 5 with a splice acceptor located upstream from the β geo gene which can be used for both LacZ staining and antibiotic selection. The
10 fusion gene possesses its own initiator methionine and an internal ribosomal entry site (IRES) for efficient translation initiation. In addition, the PGK promoter and puromycin-splice donor sequences have been flanked by lox P recombination sites. This allows for the ability to both
15 remove and introduce sequences at the integration site and is of potential value with regard to the manipulation of regions proximal to trapped target genes (Barinaga, Science 265:26-8, 1994). While this particular vector includes lox P recombination sites, the present invention is in no way
20 limited to the use of this specific recombination site (Akagi et al., Nucleic Acids Res 25:1766-73, 1997).

Another very important use of recombinases is to produce mutations that can be made tissue-specific and/or inducible. In the presently described vectors, the SA β geo or SAIRES β geo
25 component provides the mutagenic function by "trapping" the normal splicing from preceding exons. If the SA β geo is flanked by inverted loxP, frt, or any other recombinase sites, the addition of the corresponding recombinase results in the flipping of the SA β geo sequence so that it no longer
30 prevents the normal splicing of the cellular gene into which it is integrated. To make a gene trap tissue-specific or inducible one could produce the trap with SA β geo in the reverse orientation and then provide recombinase activity only at the time and place where one wishes to remove the
35 gene function. The use of tissue-specific or inducible recombinase constructs allows one to choose when and where one removes, or activates, the function of the targeted gene.

One method for practicing the inducible forms of recombinase mediated gene expression involves the use of vectors that use inducible or tissue specific promoter/operator elements to express the desired recombinase activity. The inducible expression elements are preferably operatively positioned to allow the inducible control or activation of expression of the desired recombinase activity. Examples of such inducible promoters or control elements include, but are not limited to, tetracycline, 10 metallothionine, ecdysone, and other steroid-responsive promoters, rapamycin responsive promoters, and the like (No et al., Proc Natl Acad Sci USA 93:3345-51, 1996; Furth et al., Proc Natl Acad Sci USA 91:9302-6, 1994). Additional control elements that can be used include promoters requiring 15 specific transcription factors such as viral, particularly HIV, promoters. Vectors incorporating such promoters would only express recombinase activity in cells that express the necessary transcription factors.

The incorporation of recombinase sites into the gene 20 trapping vectors highlights the value of using the described gene trap vectors to deliver specific DNA sequence elements throughout the genome. Although a variety of vectors are available for placing sequences into the genome, the presently described vectors facilitate both the insertion of 25 the specific elements, and the subsequent identification of where sequence has inserted into the cellular chromosome. Additionally, the presently described vectors may be used to place recombinase recognition sites throughout the genome. The recombinase recognition sites could then be used to 30 either remove or insert specific DNA sequences at predetermined locations.

Moreover, the described gene trap vectors can also be used to insert regulatory elements throughout the genome. Recent work has identified a number of inducible or 35 repressible systems that function in the mouse. These include the rapamycin, tetracycline, ecdysone, glucocorticoid, and heavy metal inducible systems. These

systems typically rely on placing DNA elements in or near a promoter. An inducible or repressible transcription factor that can identify and bind to the DNA element may also be engineered into the cells. The transcription factor will
5 specifically bind to the DNA element in either the presence or absence of a ligand that binds to the transcription factor and, depending on the structure of the transcription factor, it will either induce or repress the expression of the cellular gene into which the DNA elements have been inserted.
10 The ability to place these inducible or repressible elements throughout the genome would increase the value of the library by adding the potential to regulate the expression of the trapped gene.

The vectors described also have important applications
15 for the overexpression of genes or portions of genes to select for phenotypic effects. Currently, overexpression of cDNA libraries to look for genes or parts of genes with specific functions is a common practice. One example would be to overexpress genes or portions of genes to look for
20 expression that causes loss of contact inhibition for cell growth as determined by growth in soft agar. This would allow the identification of genes or portions of genes that can act as oncogenes. Simple modifications of VICTR 20 would allow it to be used for these applications. For example, the
25 addition of an internal ribosome entry site (IRES) 3' to the puromycin selectable marker and before the SD sequence, would result in the overexpression of sequences from the trapped downstream exons. In addition, the IRES could be modified by, for example, the addition of one or two nucleotides such
30 that there could be 3 basic vectors that would allow expression of trapped exons in all three reading frames. In this way, genes could be trapped throughout the genome resulting in overexpression of genes, or portions thereof, to examine the cellular function of the trapped genes. This
35 identification of function could be done by selecting for the function of interest (i.e., growth in soft agar could result from the overexpression of potentially oncogenic genes).

This technique would allow for the screening or selection of large numbers of genes, or portions thereof, by overexpressing the genes and identifying cells displaying the phenotypes of interest. Additional assays could, for example, identify candidate tumor suppressor genes based on their ability, when overexpressed, to prevent growth in soft agar.

Given the fact that expression pattern information can provide insight into the possible functions of genes mutated by the current methods, another LTR vector, VICTR 6, has been constructed in a manner similar to VICTR 5 except that the terminal exon has been replaced with either a gene coding for β -galactosidase (β gal) or a fusion between β -gal and neomycin phosphotransferase (β geo), each proceeded by a splice acceptor and followed by a polyadenylation signal.

Endogenous gene expression and splicing of these markers into cellular transcripts and translation into fusion proteins will allow for increased mutagenicity as well as the delineation of expression through Lac Z staining.

An additional vector, VICTR 12, incorporates two separate selectable markers for the analysis of both integration sites and trapped genes. One selectable marker (e.g. puro) is similar to that for VICTRs 3 through 5 in that it contains a promoter element at its 5' end and a splice donor sequence 3'. This gene cassette is located in the LTRs of the retroviral vector. The other marker (neo) also contains a promoter element but has a polyadenylation signal present at the 3' end of the coding sequence and is positioned between the viral LTRs. Both selectable markers contain an initiator ATG for proper translation. The design of VICTR 12 allows for the assessment of absolute titer as assayed by the number of colonies resistant to antibiotic selection for the constitutively expressed marker possessing a polyadenylation signal. This titer can then be compared to that observed for gene-trapping and stable expression of the resistance marker flanked at its 3' end by a splice donor. These numbers are important for the calculation of gene

trapping frequency in the context of both nonspecific binding by retroviral integrase and directed binding by chimeric integrase fusions. In addition, it provides an option to focus on the actual integration sites through infection and selection for the marker containing the polyadenylation signal. This eliminates the need for the fusion protein binding to occur upstream and in the proximity of the target gene. Theoretically, any transcription factor binding sites present within the genome are targets for proximal integration and subsequent antibiotic resistance. Analysis of sequences flanking the LTRs of the retroviral vector should reveal canonical factor binding sites. In addition, by including the promoter/splice donor design of VICTR 3, gene-trapping abilities are retained in VICTR 12.

VICTR A is a vector which does not contain gene trapping constructs but rather a selectable marker possessing all of the required entities for constitutive expression including, but not limited to, a promoter element capable of driving expression in eukaryotic cells and a polyadenylation and transcriptional terminal signal. Similar to VICTR 12, downstream gene trapping is not necessary for successful selection using VICTR A. This vector is intended solely to select for successful integrations and serves as a control for the identification of transcription factor binding sites flanking the integrant as mentioned above.

Finally, VICTR B is similar to VICTR A in that it comprises a constitutively expressed selectable marker, but it also contains the bacterial β -lactamase ampicillin resistance selectable marker and a ColE1 origin of replication. These entities allow for the rapid cloning of sequences flanking the long terminal repeats through restriction digestion of genomic DNA from infected cells and ligation to form plasmid molecules which can be rescued by bacterial transformation, and subsequently sequenced. This vector allows for the rapid analysis of cellular sequences that contain putative binding sites for the transcription factor of interest.

Other vector designs contemplated by the present invention are engineered to include an inducible regulatory elements such as tetracycline, ecdysone, and other steroid-responsive promoters (No et al., Proc Natl Acad Sci USA 5 93:3345-51, 1996; Furth et al., Proc Natl Acad Sci USA 91:9302-6, 1994). These elements are operatively positioned to allow the inducible control of expression of either the selectable marker or endogenous genes proximal to site of integration. Such inducibility provides a unique tool for 10 the regulation of target gene expression.

All of the gene trap vectors of the VICTR series, with the exception of VICTRs A and B, are designed to form a fusion transcript between vector encoded sequence and the trapped target gene. All of the flanking exons may be 15 sequenced according to the methods described in the following section. To facilitate sequencing, specific sequences are engineered onto the ends of the selectable marker (e.g., puromycin coding region). Examples of such sequences include, but are not limited to unique sequences for priming 20 PCR, and sequences complementary to standard M13 sequencing primers. Additionally, stop codons are added in all three reading frames to ensure that no anomalous fusion proteins are produced. All of the unique 3' primer sequences are immediately followed by a synthetic 9 base pair splice donor 25 sequence. This keeps the size of the exon comprising the selectable marker at a minimum to ensure proper splicing, and positions the amplification and sequencing primers immediately adjacent to the flanking trapped exons to be sequenced as part of the generation of the collection of 30 cells representing mutated transcription factor targets.

Since a cryptic splice donor sequence is found in the inverted LTRs, this cryptic splice donor sequence has been removed from the VICTR vectors by site specific mutagenesis. It was deemed necessary to remove this splice donor so that 35 it would not affect trapping associated splicing events.

When any members of the VICTR series are packaged into infectious virus, the direction of transcription of the

selectable marker is opposite to that of the direction of the normal transcription of the retrovirus. The reason for this organization is that the regulatory elements such as the polyadenylation signal, the splice sites and the promoter elements found in the various members of the VICTR series can interfere with the transcription of the retroviral genome in the packaging cell line. This potential interference may significantly reduce retroviral titers.

Although specific gene trapping vectors have been discussed at length above, the invention is by no means to be limited to such vectors. Several other types of vectors that may also be used to incorporate relatively small engineered exons into a target cell transcripts include, but are not limited to, adenoviral vectors, adenoassociated virus vectors, SV40 based vectors, and papilloma virus vectors. Additionally, DNA vectors may be directly transferred into the target cells using any of a variety of biochemical or physical means such as lipofection, chemical transfection, retrotransposition, electroporation, and the like.

Although, the use of specific selectable markers has been disclosed and discussed herein, the present invention is in no way limited to the specifically disclosed markers. Additional markers (and associated antibiotics) that are suitable for either positive or negative selection of eukaryotic cells are disclosed, *inter alia*, in Sambrook *et al.* (1989) Molecular Cloning Vols. I-III, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, and Current Protocols in Molecular Biology (1989) John Wiley & Sons, all Vols. and periodic updates thereof, as well as Table I of U.S. Patent No. 5,464,764 issued November 7, 1995, the entirety of which is herein incorporated by reference. Any of the disclosed markers, as well as others known in the art, may be used to practice the present invention.

5.2. The Analysis of Mutated Genes and Transcripts

The presently described invention allows for large-scale genetic analysis of the genomes of any organism for which

there exists cultured cell lines. The Library may be constructed from any type of cell that can be transfected by standard techniques or infected with recombinant retroviral vectors.

5 Where mouse ES cells are used, then the Library becomes a genetic tool able to completely represent mutations in essentially every gene of the mouse genome. Since ES cells can be injected back into a blastocyst and become incorporated into normal development and ultimately the germ
10 line, the cells of the Library effectively represent a complete panel of mutant transgenic mouse strains (see generally, U.S. Patent No. 5,464,764 issued November 7, 1995, herein incorporated by reference).

A similar methodology may be used to construct virtually
15 any non-human transgenic animal (or animal capable of being rendered transgenic). Such nonhuman transgenic animals may include, for example, transgenic pigs, transgenic rats, transgenic rabbits, transgenic cattle, transgenic goats, and other transgenic animal species, particularly mammalian
20 species, known in the art. Additionally, bovine, ovine, and porcine species, other members of the rodent family, e.g. rat, as well as rabbit and guinea pig and non-human primates, such as chimpanzee, may be used to practice the present invention.

25 Transgenic animals produced using the presently described library and/or vectors are useful for the study of basic biological processes and diseases including, but not limited to, aging, cancer, autoimmune disease, immune disorders, alopecia, glandular disorders, inflammatory
30 disorders, diabetes, arthritis, high blood pressure, atherosclerosis, cardiovascular disease, pulmonary disease, degenerative diseases of the neural or skeletal systems, Alzheimer's disease, Parkinson's disease, asthma, developmental disorders or abnormalities, infertility,
35 epithelial ulcerations, and microbial pathogenesis (a relatively comprehensive review of such pathogens is provided, *inter alia*, in Mandell et al., 1990, "Principles

and Practice of Infectious Disease" 3rd. ed., Churchill Livingstone Inc., New York, N.Y. 10036, herein incorporated by reference). As such, the described animals and cells are particularly useful for the practice of functional genomics.

5

5.2.1. Constructing a Library of Individually Mutated Cell Clones

The vectors described in the previous section were used to infect (or transfect) cells in culture, for example, mouse embryonic stem (ES) cells. Gene trap
10 insertions were initially identified by antibiotic resistance (e.g., puromycin). Individual clones (colonies) were moved from a culture dish to individual wells of a multi-welled tissue culture plate (e.g. one with 96 wells). From this
15 platform, the clones were be duplicated for storage and subsequent analysis. Each multi-well plate of clones was then processed by molecular biological techniques described in the following section in order to derive sequence of the gene that has been mutated. This entire process is presented
20 schematically in Figure 4 (described below).

5.2.2. Identifying and Sequencing the Tagged Genes in the Library.

The relevant nucleic acid (and derived amino acid sequence information) will largely be obtained using
25 PCR-based techniques that rely on knowing part of the sequence of the fusion transcripts (see generally, Frohman et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85(23):8998-9000, and U.S. Patents Nos. 4,683,195 to Saiki et al., and 4,683,202 to Mullis, which are herein incorporated by
30 reference). Typically, such sequences are encoded by the foreign exon containing the selectable marker. The procedure is represented schematically in Figure 2 (3' RACE). Although each step of the procedure may be done manually, the procedure is also designed to be carried out using robots
35 that can deliver reagents to multi well culture plates (e.g., but not limited to, 96-well plates).

The first step generates single stranded complementary DNA which is used in the PCR amplification reaction (Figure 2). The RNA substrate for cDNA synthesis may either be total cellular RNA or an mRNA fraction; preferably the latter.

5 mRNA was isolated from cells directly in the wells of the tissue culture dish. The cells were lysed and mRNA was bound by the complementary binding of the poly-adenylate tail to a poly-thymidine-associated solid matrix. The bound mRNA was washed several times and the reagents for the reverse

10 transcription (RT) reaction were added. cDNA synthesis in the RT reaction was initiated at random positions along the message by the binding of a random sequence primer (RS). This RS primer has approximately 6-9 random nucleotides at the 3' end to bind sites in the mRNA to prime cDNA synthesis,

15 and a 5' tail sequence of known composition to act as an anchor for PCR amplification in the next step. There is therefore no specificity for the trapped message in the RT step. Alternatively, a poly-dT primer appended with the specific sequences for the PCR may be used. Synthesis of the

20 first strand of the cDNA initiates at the end of each trapped gene. At this point in the procedure, the bound mRNA may be stored (at between about -70° C and about 4° C) and reused multiple times. Such storage is a valuable feature where one subsequently desires to analyze individual clones in more

25 detail. The bound mRNA may also be used to clone the entire transcript using PCR-based protocols.

Specificity for the trapped, fusion transcript is introduced in the next step, PCR amplification. The primers for this reaction are complementary to the anchor sequence of

30 the RS primer and to the selectable marker. Double stranded fragments between a fixed point in the selectable marker gene and various points downstream in the appended transcript sequence are amplified. It is these fragments which will become the substrates for the sequencing reaction. The

35 various end-points along the transcript sequence were determined by the binding of the random primer during the RT reaction. These PCR products were diluted into the

sequencing reaction mix, denatured and sequenced using a primer specific for the splice donor sequences of the gene trap exon. Although, standard radioactively labeled nucleotides may be used in the sequencing reactions, 5 sequences will typically be determined using standard dye terminator sequencing in conjunction with automated sequencers (e.g., ABI sequencers and the like).

Several fragments of various sizes may serve as substrates for the sequencing reactions. This is not a 10 problem since the sequencing reaction proceeds from a fixed point as defined by a specific primer sequence. Typically, approximately 200 nucleotides of sequence were obtained for each trapped transcript. For the PCR fragments that are shorter than this, the sequencing reaction simply 'falls off' 15 the end. Sequences further 3' were then covered by the longer fragments amplified during PCR. One problem is presented by the anchor sequences 'S' derived from the RS primer. When these are encountered during the sequencing of smaller fragments, they register as anomalous dye signals on 20 the sequencing gels. To circumvent this potential problem, a restriction enzyme recognition site is included in the S sequence. Digestion of the double stranded PCR products with this enzyme prior to sequencing eliminates the heterologous S sequences.

25

5.2.3. Identifying the Tagged Genes by Chromosomal Location

Any individually tagged gene may also be identified by PCR using chromosomal DNA as the template. To 30 find an individual clone of interest in the Library arrayed as described above, genomic DNA is isolated from the pooled clones of ES cells as presented in Figure 3. One primer for the PCR is anchored in the gene trap vector, e.g., a puro exon-specific oligonucleotide. The other primer is located 35 in the genomic DNA of interest. This genomic DNA primer may consist of either (1) DNA sequence that corresponds to the coding region of the gene of interest, or (2) DNA sequence

from the locus or the gene of interest. In the first case, the only way that the two primers used may be juxtaposed to give a positive PCR results (e.g., the correct size double-stranded DNA product) is if the gene trap vector has inserted 5 into the gene of interest. Additionally, degenerate primers may be used, to identify and isolate related genes of interest. In the second case, the only way that the two primers used may be juxtaposed to provide the desired PCR result is if the gene trap vector has inserted into the 10 region of interest that contains the primer for the known marker.

For example, if one wishes to obtain ES cell clones from the library that contain mutated genes located in a certain chromosomal position, PCR primers are designed that 15 correspond to the puro gene (the puro-anchored primer) and a primer that corresponds to a marker known to be located in the region of interest. Several different combinations of marker primers and primers that are located in the region of interest may also be used to obtain optimum results. In this 20 manner, the mutated genes are identified by virtue of their location relative to sets of known markers. Genes in a particular chromosomal region of interest could therefore be identified. The marker primers could also be designed correspond to sequences of known genes in order to screen for 25 mutations in particular genes by PCR on genomic DNA templates. While this method is likely to be less informative than the RT-PCR strategy described below, this technique would be useful as a alternative strategy to identify mutations in known genes. In addition, primers that 30 correspond to sequence of known genes could be used in PCR reactions with marker-specific primers in order to identify ES cell clones that contain mutations in genes proximal to the known genes. The sensitivity of detection is adequate to find such events when positive clones are subsequently 35 identified as described below in the RT-PCR strategy.

5.3. A Sequence Database Identifies Genes Mutated in the Library.

Using the procedures described above, approximately 200 to about 600 bases of sequence from the cellular exons appended to the selectable marker exon (e.g., *puro* exon in VICTR vectors) may be identified. These sequences provide a means to identify and catalogue the genes mutated in each clone of the Library. Such a database provides both an index for the presently disclosed libraries, and a resource for discovering novel genes. Alternatively, various comparisons can be made between the Library database sequences and any other sequence database as would be familiar to those practiced in the art.

The novel utility of the Library lies in the ability for a person to search the Library database for a gene of interest based upon some knowledge of the nucleic acid or amino acid sequence. Once a sequence is identified, the specific clone in the Library can be accessed and used to study gene function. This is accomplished by studying the effects of the mutation both *in vitro* and *in vivo*. For example, cell culture systems and animal models (i.e., transgenic animals) may be directly generated from the cells found in the Library as will be familiar to those practiced in the art.

Additionally, the sequence information may be used to generate a highly specific probe for isolating both genomic clones from existing data bases, as well as a full length cDNA. Additionally, the probe may be used to isolate the homologous gene from sufficiently related species, including humans. Once isolated, the gene may be over expressed, or used to generate a targeted knock-out vector that may be used to generate cells and animals that are homozygous for the mutation of interest. Such animals and cells are deemed to be particularly useful as disease models (i.e., cancer, genetic abnormalities, AIDS, etc.), for developmental study, to assay for toxin susceptibility or the efficacy of therapeutic agents, and as hosts for gene delivery and

therapy experiments (e.g., experiments designed to correct a specific genetic defect *in vivo*).

5.4. Accessing Clones in the Library by a Pooling and Screening Procedure.

5 An alternative method of accessing individual clones is by searching the Library database for sequences in order to isolate a clone of interest from pools of library clones. The Library may be arrayed either as single clones, each with
10 different insertions, or as sets of pooled clones. That is, as many clones as will represent insertions into essentially every gene in the genome are grown in sets of a defined number. For example, 100,000 clones can be arrayed in 2,000 sets of 50 clones. This can be accomplished by titrating the
15 number of VICTR retroviral particles added to each well of 96-well tissue culture plates. Two thousand clones will fit on approximately 20 such plates. The number of clones may be dictated by the estimated number of genes in the genome of the cells being used. For example, there are approximately
20 100,000 genes in the genome of mouse ES cells. Therefore, a Library of mutations in essentially every gene in the mouse genome may be arrayed onto 20 96-well plates.

To find an individual clone of interest in the Library arrayed in this manner, reverse transcription-polymerase
25 chain reactions (RT-PCR) are performed on mRNA isolated from pooled clones as presented in Figure 4. One primer for RT-PCR is anchored in the gene trap vector, i.e. a *puro* exon-specific oligonucleotide. The other primer is located in the cDNA sequence of a gene of interest. The only way that these
30 two sequences can be juxtaposed to give a positive RT-PCR result (i.e. double stranded DNA fragment visible by agarose gel electrophoresis, as will be familiar to anyone practiced in the art) is by being present in a transcript from a gene trap event occurring in the gene of interest.

35 For example, if one wishes to obtain an ES cell clone with a mutation in the *p53* gene, PCR primers are designed that correspond to the *puro* and *p53* genes. If a VICTR

trapping vector integrates into the p53 locus and results in the formation of a fusion mRNA, this mRNA may be detected by RT-PCR using these specifically designed primer pairs. The sensitivity of detection is adequate to find such an event
5 when positive cells are mixed with a large background of negative cells. The individual positive clones are subsequently identified by first locating the pool of 50 clones in which it resides. This process is described in Figure 5. The positive pool, once identified, is
10 subsequently plated at limiting dilution (approximately 0.3 cells/well) such that individual clones may be isolated. To find the one positive event in 50 clones represented by this pool, individual clones are isolated and arrayed on a 96-well plate. By pooling in columns and rows, the positive well
15 containing the positive clone can be identified with relatively few RT-PCR reactions.

In addition to RT-PCR, the pools may be screened by hybridization techniques (see generally Sambrook et al., 1989, Molecular Cloning: H Laboratory Manual 2nd edition,
20 Cold Spring Harbor Press, Cold Spring Harbor, and Current Protocols in Molecular Biology, 1995, Ausubel et al. eds., John Wiley and Sons). Specific PCR fragments are generated from the mutated genes essentially as described above for the sequencing protocols of the individual clones (first-strand
25 synthesis using RT primed by a random or oligo dT primer that is appended to a specific primer binding site). The gene trap DNA is amplified from the primer sets in the puro gene and the specific sequences appended to the RT primer. If this were done with pools, the resulting pooled set of
30 amplified DNA fragments could be arrayed on membranes and probed by radioactive, or chemically or enzymatically labeled, hybridization probes specific for a gene of interest. A positive radioactive result indicates that the gene of interest has been mutated in one of the clones of the
35 positively-labeled pool. The individual positive clone is subsequently identified by PCR or hybridization essentially as outlined above.

Alternatively, a similar strategy may be used to identify the clone of interest from multiple plates, or any scheme where a two or three dimensional array (e.g., columns and rows) of individual clones are pooled by row or by column. For example, 96 well plates of individual clones may be arranged adjacent to each other to provide a larger (or virtual/figurative) two dimensional grid (e.g., four plates may be arranged to provide a net 16x24 grid), and the various rows and columns of the larger grid may be pooled to achieve substantially the same result.

Similarly, plates may simply be stacked, literally or figuratively, or arranged into a larger grid and stacked to provide three dimensional arrays of individual clones. Representative pools from all three planes of the three dimensional grid may then be analyzed, and the three positive pools/planes may be aligned to identify the desired clone. For example, ten 96 well plates may be screened by pooling the respective rows and columns from each plate (a total of 20 pools) as well as pooling all of the clones on each specific plate (10 additional pools). Using this method, one may effectively screen 960 clones by performing PCR on only 30 pooled samples.

The example provided below is merely illustrative of the subject invention. Given the level of skill in the art, one may be expected to modify any of the above or following disclosure to produce insubstantial differences from the specifically described features of the present invention. As such, the following example is provided solely by way of illustration and is not included for the purpose of limiting the invention in any way whatsoever.

6.0. EXAMPLES

6.1. Use of VICTR Series Vectors to Construct a Mouse ES cell Gene Trap Library

VICTR 3 was used to gather a set of gene trap clones. A plasmid containing the VICTR 3 cassette was constructed by conventional cloning techniques and designed to employ the

features described above. Namely, the cassette contained a
PGK promoter directing transcription of an exon that encodes
the puro marker and ends in a canonical splice donor
sequence. At the end of the puromycin exon, sequences were
5 added as described that allow for the annealing of two nested
PCR and sequencing primers. The vector backbone was based on
pBluescript KS+ from Stratagene Corporation.

The plasmid construct linearized by digestion with Sca I
which cuts at a unique site in the plasmid backbone. The
10 plasmid was then transfected into the mouse ES cell line
AB2.2 by electroporation using a BioRad Genepulser apparatus.
After the cells were allowed to recover, gene trap clones
were selected by adding puromycin to the medium at a final
concentration of 3 μ g/mL. Positive clones were allowed to
15 grow under selection for approximately 10 days before being
removed and cultured separately for storage and to determine
the sequence of the disrupted gene.

Total RNA was isolated from an aliquot of cells from
each of 18 gene trap clones chosen for study. Five
20 micrograms of this RNA was used in a first strand cDNA
synthesis reaction using the "RS" primer. This primer has
unique sequences (for subsequent PCR) on its 5' end and nine
random nucleotides or nine T (thymidine) residues on its 3'
end. Reaction products from the first strand synthesis were
25 added directly to a PCR with outer primers specific for the
engineered sequences of puromycin and the "RS" primer. After
amplification, an aliquot of reaction products were subject
to a second round of amplification using primers internal, or
nested, relative to the first set of PCR primers. This
30 second amplification provided more reaction product for
sequencing and also provided increased specificity for the
specifically gene trapped DNA.

The products of the nested PCR were visualized by
agarose gel electrophoresis, and seventeen of the eighteen
35 clones provided at least one band that was visible on the gel
with ethidium bromide staining. Most gave only a single band
which is an advantage in that a single band is generally

easier to sequence. The PCR products were sequenced directly after excess PCR primers and nucleotides were removed by filtration in a spin column (Centricon-100, Amicon). DNA was added directly to dye terminator sequencing reactions
5 (purchased from ABI) using the standard M13 forward primer a region for which was built into the end of the puro exon in all of the PCR fragments. Thirteen of the seventeen clones that gave a band after the PCR provided readable sequence. The minimum number of readable nucleotides was 207 and some
10 of the clones provided over 500 nucleotides of useful sequence.

Sample data from this set of clones is presented in Figure 6. Only a portion of sequence (nucleotide or putative amino acid) for 9 Library clones obtained by the methods
15 described in this invention are presented. Under each sequence fragment in the figure is aligned a homologous sequence that was identified using the BLAST (basic local alignment search tool) search algorithm (Altschul et al., 1990, J. Mol. Biol. 215:403-410).

20 In addition to known sequences, many new genes were also identified. Each of these sequences is labeled "OST" for "Omnibank Sequence Tags." OMNIBANK™ shall be the trademark name for the Libraries generated using the disclosed technology.

25 These data demonstrate that the VICTR series vectors may efficiently trap genes, and that the procedures used to obtain sequence are reliable. With simple optimization of each step, it is presently possible to mutate every gene in a given population of cells, and obtain sequence from each of
30 these mutated genes. The sample data provided in this example represents a small fraction of an entire Library. By simply performing the same procedures on a larger scale (with automation) a Library may be constructed that collectively comprises and indexes mutations in essentially every gene in
35 the genome of the target cell.

Additional studies have used both VICTR 3 and VICTR 20. Like VICTR 3, VICTR 20 is exemplary of a family of vectors

that incorporate two main functional units: a sequence acquisition component having a strong promoter element (phosphoglycerate kinase 1) active in ES cells that is fused to the puromycin resistance gene coding sequence which lacks
5 a polyadenylation sequence but is followed by a synthetic consensus splice donor sequence (PGKpuroSD); and 2) a mutagenic component that incorporates a splice acceptor sequence fused to a selectable, colorimetric marker gene and followed by a polyadenylation sequence (for example, SA β geopA
10 or SAIRES β geopA). Also like VICTR 3, stop codons have been engineered into all three reading frames in the region between the 3' end of the selectable marker and the splice donor site. A diagrammatic description of structure and functions of VICTRs 3 and 20 is provided in Figure 7.

15 When VICTRs 3 and 20 were used in the commercial scale application of the presently disclosed invention, over 3,000 mutagenized ES cell clones were rapidly engineered and obtained. Sequence analysis obtained from these clones has identified a wide variety of both previously identified and
20 novel sequences. A representative sampling of previously known genes that were identified using the presently described methods is provided in Figure 8. The power of the presently described invention as a genomics resource becomes apparent when one considers that the genes listed in Figure 8
25 were obtained and identified in less than a year whereas the references associated with the identification of the known genes span a period of roughly two decades. More importantly, the majority of the sequences thus far identified are novel, and, because of the functional aspects
30 of the presently described ES cell system, the cellular and developmental functions of these novel sequences can be rapidly established.

7.0. Reference to Microorganism Deposits

35 The following plasmids have been deposited at the American Type Culture Collection (ATCC), Rockville, MD, USA, under the terms of the Budapest Treaty on the International

Recognition of the Deposit of Microorganisms for the Purposes
of Patent Procedure and Regulations thereunder (Budapest
Treaty) and are thus maintained and made available according
to the terms of the Budapest Treaty. Availability of such
5 plasmids is not to be construed as a license to practice the
invention in contravention of the rights granted under the
authority of any government in accordance with its patent
laws.

The deposited cultures have been assigned the indicated
10 ATCC deposit numbers:

	<u>Plasmid</u>	<u>ATCC No.</u>
	plex	97748
	pExonII	97749
	ppuro7	97750
	ppuro5	97751
15	ppuro11	97752
	ppuro10	97753

All publications and patents mentioned in the above
specification are herein incorporated by reference. Various
modifications and variations of the described method and
system of the invention will be apparent to those skilled in
20 the art without departing from the scope and spirit of the
invention. Although the invention has been described in
connection with specific preferred embodiments, it should be
understood that the invention as claimed should not be unduly
limited to such specific embodiments. Indeed, various
25 modifications of the above-described modes for carrying out
the invention which are obvious to those skilled in the field
of molecular biology or related fields are intended to be
within the scope of the following claims.

30

35

CLAIMS

What is claimed is:

1. A library of cultured eucaryotic cells made by a process comprising the steps of:
 - 5 a) treating a first group of cells to stably integrate a first vector that mediates the splicing of a foreign exon internal to a cellular transcript;
 - b) treating a second group of cells to stably integrate a second vector that mediates the splicing of a foreign exon
10 5' to an exon of a cellular transcript; and
 - c) selecting for transduced cells that express the products encoded by the foreign exons.
2. The library of claim 1 wherein said treating is
15 transfection.
3. The library of claim 1 wherein said treating is by infection.
- 20 4. The library of claim 1 wherein said treating is by retrotransposition.
5. The library of any one of claims 1 through 4 wherein said cells are animal cells.
25
6. The library of claim 5 wherein said animal is mammalian.
7. The library of claim 6 wherein said cells are rodent
30 cells.
8. The use of a mutated cell from a library according to claim 6 to generate a non-human transgenic animal.
- 35 9. A vector for replacing the 3' end of an animal cell transcript with a foreign exon, comprising:
 - a) a selectable marker;

- b) a splice acceptor site operatively positioned 5' to the initiation codon of said selectable marker;
- c) a polyadenylation site operatively positioned 3' to said selectable marker;
- 5 d) said vector not comprising a promoter element operatively positioned 5' of the coding region of said selectable marker; and
- e) said vector not comprising a splice donor sequence operatively positioned between the 3' end of the
- 10 coding region of said selectable marker and said polyadenylation site.

10. A vector for inserting foreign mutagenic polynucleotide sequence internal to animal cell transcripts, comprising:

- a) a foreign exon;
- b) a splice acceptor sequence operatively positioned 5' to the foreign exon;
- c) a splice donor site operatively positioned 3' to said foreign exon;
- 20 d) a sequence comprising a nested set of stop codons in each of the three reading frames located between the 3' end of said foreign exon and said splice donor site;
- 25 e) said vector not comprising a polyadenylation site operatively positioned 3' to said foreign exon; and
- f) said vector not comprising a promoter element operatively positioned 5' to the coding region of said foreign exon.

30

11. A vector for attaching a foreign exon upstream from the 3' end of an animal cell transcript, comprising:

- a) a selectable marker;
- b) a promoter element operatively positioned 5' to said selectable marker;
- 35 c) a splice donor site operatively positioned 3' to said selectable marker; and

- d) said vector not comprising a transcription terminator or polyadenylation site operatively positioned relative to the coding region of said selectable marker; and
- 5 e) said vector not comprising a splice acceptor site operatively positioned between said promoter element and the initiation codon of said selectable marker.
- 10 12. The vector of claim 11 wherein said vector additionally comprises a foreign mutagenic polynucleotide sequence located upstream from said promoter.
- 15 13. The vector of claim 12 wherein said vector additionally comprises a splice acceptor operatively positioned upstream from said foreign mutagenic polynucleotide sequence.
- 20 14. The vector of claim 13 wherein said foreign mutagenic polynucleotide sequence comprises a polyadenylation site.
- 25 15. The vector of claim 14, wherein said foreign mutagenic polynucleotide sequence additionally comprises stop codons in all three reading frames.
- 30 16. The vector of claim 12 in which a first recombinase recognition sequence is present upstream from said promoter and a second recombinase recognition sequence is present downstream from said promoter.
17. The vector of any one of claims 9, 10, or 11 wherein said vector is a viral vector.
- 35 18. The vector of claim 17 wherein said viral vector is a retroviral vector.

19. The use of a vector according to claim 9 to produce a library of mutated animal cells.

20. The use of a vector according to claim 10 to produce mutated animal cells.

21. The use of a vector according to claim 11 to produce mutated animal cells.

10 22. The use of a vector according to claim 11 to effect homologous recombination in an animal cell.

23. A stably transduced animal cell that incorporates a vector according to claim 16.

15

24. A method of deleting a region of vector DNA from a cell according to claim 23, comprising:

- a) providing a recombinase activity to the cell; and
 - b) selecting for cells that lack the desired region of
- 20 vector DNA.

25. A method of adding a region of DNA to a cell according to claim 23, comprising:

- a) introducing the DNA to be added into the cell;
- 25 a) providing a recombinase activity to the cell; and
- b) selecting for cells that incorporate the added DNA.

26. A method of effecting the inducible expression of a desired gene, comprising:

- 30 a) providing a cell according to claim 23 with a recombinase gene that is controlled by an inducible promoter; and
- b) inducing said inducible promoter.

35 27. A method of gene discovery comprising:

- a) adding a foreign polynucleotide to a population of target cells such that the foreign

polynucleotide is inserted throughout the genomes of the target cells; and

b) activating control elements encoded by the foreign polynucleotides that activate or repress the expression of target cell genes that flank the integrated foreign polynucleotides, and identifying the regions of the target cell genome into which the foreign polynucleotides have integrated.

28. A library of cultured animal cells that stably integrate a vector according to any one of claims 10 or 11.

15

20

25

30

35

Abstract

Methods and vectors (both DNA and retroviral) are provided for the construction of a Library of mutated cells. The Library will preferably contain mutations in essentially
5 all genes present in the genome of the cells. The nature of the Library and the vectors allow for methods of screening for mutations in specific genes, and for gathering nucleotide sequence data from each mutated gene to provide a database of tagged gene sequences. Such a database provides a means to
10 access the individual mutant cell clones contained in the Library. The invention includes the described Library, methods of making the same, and vectors used to construct the Library. Methods are also provided for accessing individual parts of the Library either by sequence or by pooling and
15 screening. The invention also provides for the generation of non-human transgenic animals which are mutant for specific genes as isolated and generated from the cells of the Library.

20

25

30

35

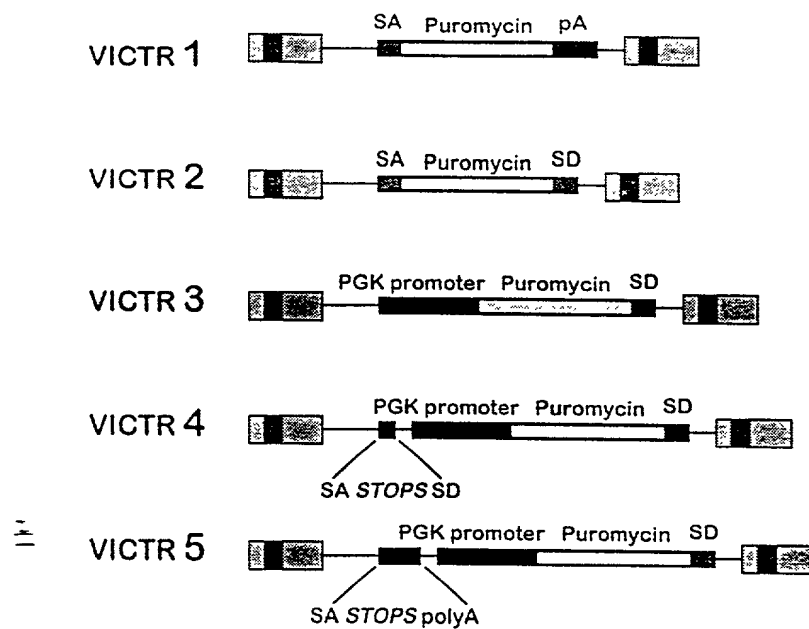


Figure 1

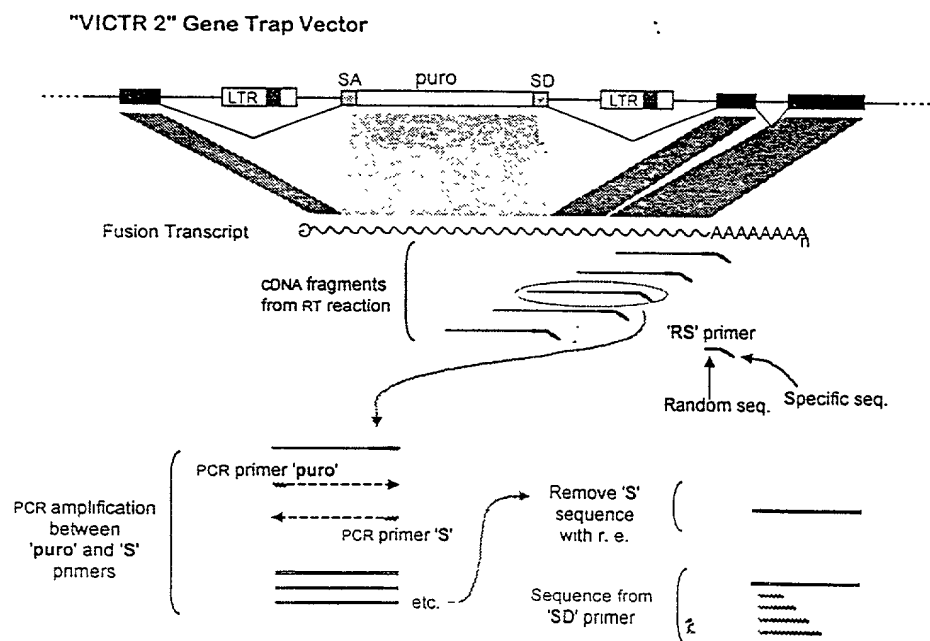


Figure 2

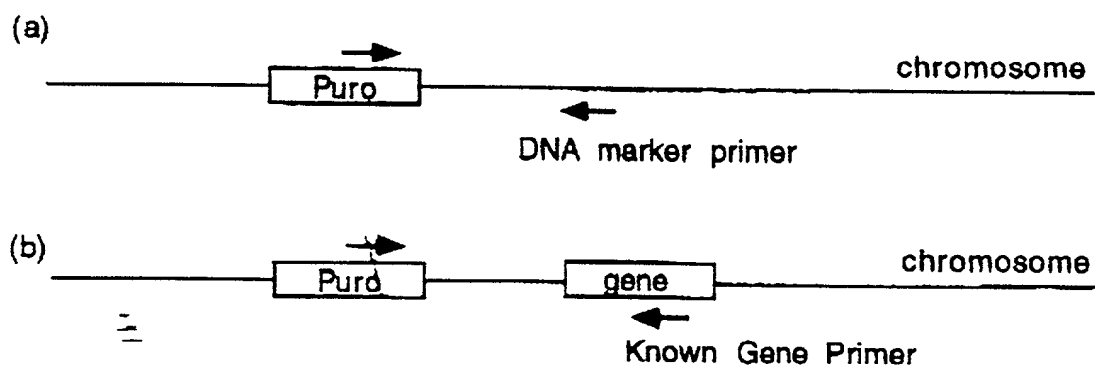


Figure 3

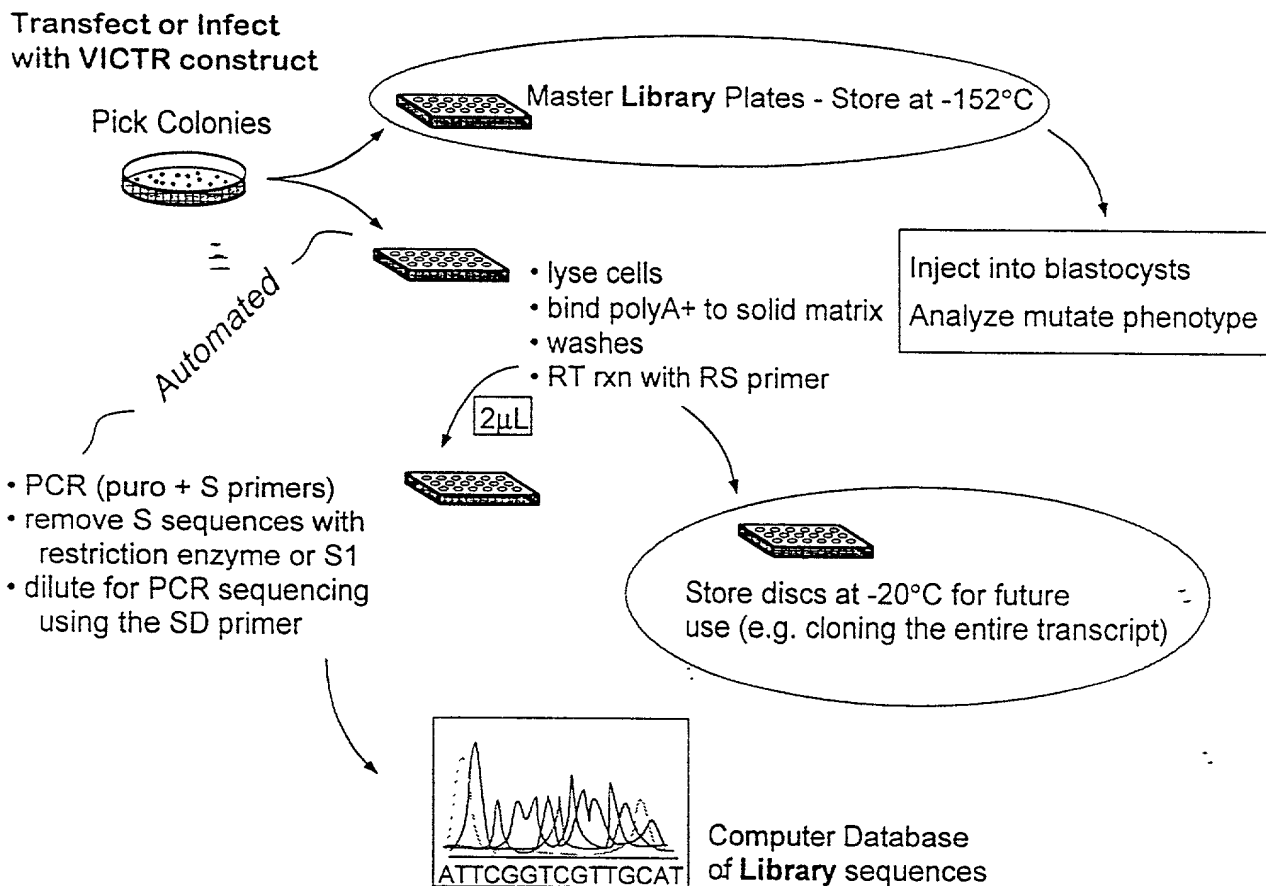
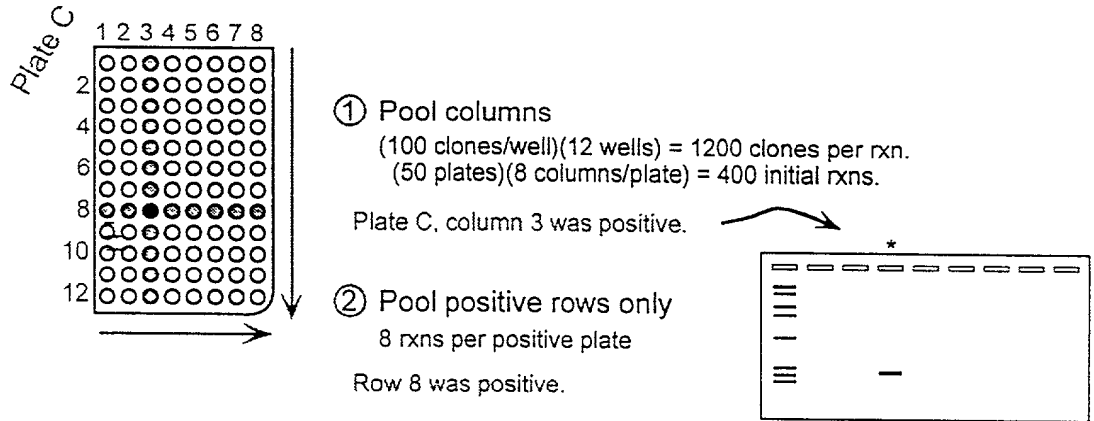


Figure 4

Identify Positive Pool

To screen all mouse genes (~100,000) with 5-fold redundancy would require about 50 plates of 96-wells (at 100 clones/well).



Identify Positive Clone

The pool on plate C, column 3, row 8 is thawed and plated as single clones:

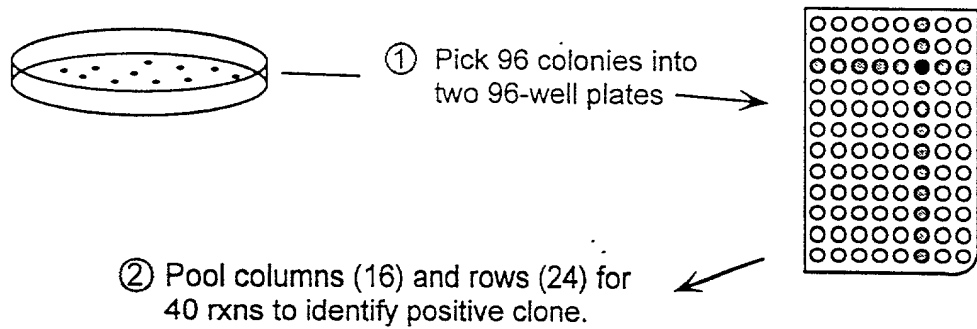


Figure 5

OST1:	248	TTTATATAATATTTAATTTGTTTTACTGGGGTATATATGTGTGAAGAGGACTTCT	302
rat GABA rho3:	1547	TTTACATAATATTTAATTTGTTTTACTGGGGTATATATGTGTGAAGAGGACTTTT	1601
OST2:	56	ACCGTTGCGGAGGCTCACGTTTCTCAGATAGTACATCAGGTGTCATCGNTGTCAGAAGGT	115
mouse TCR-ATF1:	75	ACCGTTGCGGGGCTCACGTTTCTCAGATAGTACATCAGGTGTCATCGTTATCAGAAAGT	134
OST3:	58	GIGMHAGLHERDRKTVEELFXNCKVQVLIATSTLAWGVNFPahlVliIKGTEYYDGKTRR	237
Yeast ORF G9365:	1430	GIG+HHAGL ++DR +LF K+Q+LIATSTLAWGVN PAHLVliIKGT+++D K GIGLHHAGLVQKDRSISHQLFQKNKIQLIATSTLAWGVNLPahlVliIKGTQFFDAKIEG	1489
OST4: seq. from US patent 5470724:	137	GCGCAGAAGTGGTCTGGAANTTTNTCCGCCNCCATCCAGTCTATTAAATTGTTGACNGGA	196
	166	GCGCAGAAGTGGTCTGCAACTTTATCCGCCTCCATCCAGTCTATTAAATTGTTGCCGGGA	225
OST5: mouse wnt-5A protein precursor:	108	TCWIRLGT*RXVGASLEYEYIRAS	179
		TCW++L R VG +L+ +Y A+	
	250	TCWLQLADFRKVGDALKEKYDSAA	273
OST6: human prolyl endopeptidase:	78	CTTATATGGCTACGGCGGCTTCAACATCTCCATTACACCCAACTACAGCGTGTCCAGGCT	137
	1407	CTTATATGGCTATGGCGGCTTCAACATATCCATCACACCCAACTACAGTGTTCAGGCT	1466
OST7: mouse 45S pre rRNA:	109	AAAGCATGTAGCAGTTGTAGGACACACTAGACGAGAGCACCAGATCTCATTGTGGGTGGT	168
	1604	AAAGCATGTAGCAGTTGTAGGACACACTAGACGAGAGCACCAGATCTCATTGTGGGTGGT	1663
OST8: rat MAL:	161	TGGATGCAGNCTACCACTGTGTGGCTGCCCTATTTTACCTCAGTGCCTCAGTTCTGGAAG	220
	306	TGGATGCAGCCTACCACTGTGTGGCTGCCCTGTTTTACCTCAGTGCCTCAGTCCTGGAAG	365
OST9: mouse malic enzyme:	103	ACCTGATTGTTATCCGTGGCCTGCAGAAGTCCAGAAAATACAGACCAAAGTCAACCAGTA	162
	1666	ACCTGATTGTTATCCGTGGCCTGCAGAAGTCCAGAAAATACAGACCAAAGTCAACCAGTA	1725

Figure 6

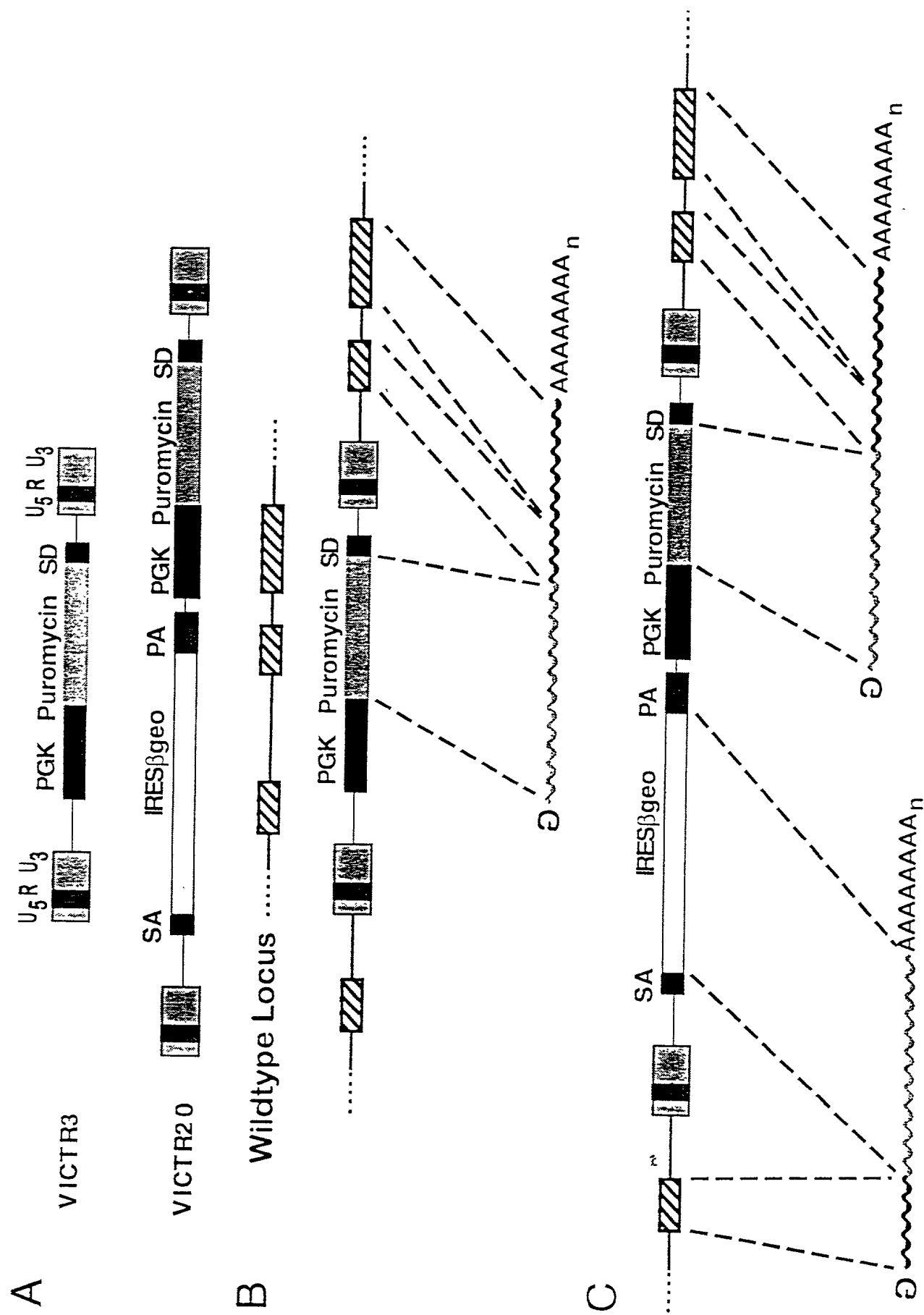


Figure 7

Accession	Gene	Species	Length	Accession	Gene	Species	Length
U01281	gb U005313	Mus musculus	1 8e-180	05T562	gb X61433	Mus musculus	7 6e-68
05T295	gb AA048390	Mus musculus	1 2e-60	05T568	gb AA007910	Mus musculus	1 5e-11
05T297	gb X71505	Mus musculus	3 0e-168	05T571	gb AA11278	Mus musculus	2 1e-147
05T300	gb H71122	Mus musculus	1 8e-203	05T572	gb AA130347	Mus musculus	1 2e-103
05T301	gb W40450	Mus musculus	2 7e-97	05T573	gb J42855	Mus musculus	4 0e-69
05T311	gb W40427	Mus musculus	3 0e-73	05T574	gb AA020459	Mus musculus	2 1e-91
05T314	gb P34710	Mus musculus	4 0e-54	05T581	gb H96552	Mus musculus	2 0e-90
05T316	gb W11499	Mus musculus	1 2e-72	05T582	gb D17695	Mus musculus	1 9e-218
05T328	gb W10861	Mus musculus	3 7e-59	05T591	gb J43126	Mus musculus	3 6e-103
05T331	gb U17698	Mus musculus	6 8e-119	05T593	gb W10777	Mus musculus	1 4e-117
05T342	gb U10120	Mus musculus	1 1e-143	05T594	gb X94616	Mus musculus	2 6e-142
05T356	gb H60456	Mus musculus	1 8e-117	05T595	gb U67137	Mus musculus	7 0e-51
05T361	gb W71360	Mus musculus	5 7e-37	05T598	gb X53476	Mus musculus	2 2e-235
05T368	gb D81662	Mus musculus	2 9e-184	05T600	gb U10494	Mus musculus	1 0e-188
05T386	gb X99946	Mus musculus	2 6e-35	05T607	gb W55702	Mus musculus	1 2e-71
05T389	gb T51727	Mus musculus	1 8e-78	05T613	gb AA184009	Mus musculus	9 8e-68
05T401	gb W29420	Mus musculus	3 1e-33	05T618	gb U11817	Mus musculus	1 5e-95
05T411	gb H48542	Mus musculus	2 0e-68	05T620	gb AA117282	Mus musculus	1 0e-78
05T418	gb G21163	Mus musculus	1 7e-84	05T623	gb AA001326	Mus musculus	5 7e-106
05T421	gb G25565	Mus musculus	6 1e-56	05T626	gb D83768	Mus musculus	1 4e-47
05T425	gb X04480	Mus musculus	8 1e-58	05T663	gb AA028410	Mus musculus	3 2e-114
05T430	gb W91937	Mus musculus	5 7e-93	05T664	gb U11027	Mus musculus	2 6e-106
05T439	gb H26756	Mus musculus	2 4e-134	05T671	gb J564860	Mus musculus	8 4e-211
05T442	gb W25938	Mus musculus	2 6e-49	05T679	gb U13516	Mus musculus	9 9e-139
05T448	gb Y107569	Mus musculus	4 1e-12	05T680	gb J20258	Mus musculus	4 2e-232
05T451	gb X95591	Mus musculus	3 1e-206	05T702	gb H78893	Mus musculus	5 7e-52
05T456	gb W75435	Mus musculus	4 6e-75	05T707	gb H15122	Mus musculus	1 2e-85
05T462	gb W13808	Mus musculus	6 8e-216	05T716	gb U67791	Mus musculus	4 5e-74

LEVA 1034 MS 03 JANUARS .5 961111

Figure 8 cont'd.

0512297	gb E07134	7 8e-34	86%	SW YAG6, 5CHHO Q09190 HYPOPHYSICAL 23.7 KD PROTEIN C1366 14 IN CHROMOSOME 1. HOMO SAPIENS H. SAPIENS PARTIAL cDNA SEQUENCE, CLONE C-52007
05T2307	gb X61399	7 0e-66	92%	MUS MUSCULUS MOUSE F502 NOVEL PROTEIN
05T2321	gb AA100747	5 1e-85	78%	HOMO SAPIENS A191104 r1 STRATAGENE CLONE 115304, HOMO SAPIENS cDNA CLONE 51203
05T2322	gb J02811	1.2e-81	92%	CLONE SP. RAG HYPOADENYLATE DEAMINASE (HMD) (HUMAN) mRNA, COMPLETE cDNA (HMD) (HUMAN) mRNA, COMPLETE cDNA
05T2346	gb AA000893	1.5e-95	94%	SPYRLOTOLAMIN 3 COMPLETE cds MUS MUSCULUS MOUSE mRNA FOR HMD (HUMAN) mRNA, COMPLETE cDNA
05T2347	gb D17653	2.5e-101	85%	CLONE 51203
05T2351	gb T80097	7.4e-73	85%	HOMO SAPIENS YD04612 r1 HOMO SAPIENS cDNA CLONE 24692 5.
05T2359	gb W30066	4.1e-133	99%	MUS MUSCULUS ME21A04 r1 SOURCE MOUSE p1NM19.5 MUS MUSCULUS cDNA CLONE 349336 5.
05T2361	gb W18873	7 5e-58	95%	MUS MUSCULUS ME04A04 r1 SOURCE MOUSE p1NM19.5 MUS MUSCULUS cDNA CLONE 317927 5.
05T2367	gb X97831	1.9e-104	85%	RATTUS NORVEGICUS R NORVEGICUS mRNA FOR CARNITINE/ACETYLCOARNINE CARRIER PROTEIN
05T2368	gb AA013837	6 1e-32	77%	MUS MUSCULUS MH24C06 r1 SOURCE MOUSE PLACENTA 4NM13.5 14.5 MUS MUSCULUS cDNA CLONE 443434 5.
05T2379	gb L10911	1 2e-105	91%	HOMO SAPIENS HOMO SAPIENS SPLICING FACTOR (CC1 4) mRNA, COMPLETE cDNA
05T2380	gb W87091	9 5e-88	90%	MUS MUSCULUS MF58D10 r1 SOURCE MOUSE EMBRYO 41B433 5.
05T2381	gb AA000090	8 1e-126	96%	MUS MUSCULUS MJ31H06 r1 SOURCE MOUSE p1NM19.5 MUS MUSCULUS cDNA CLONE 443434 5.
05T2382	gb AA013380	5.1e-126	92%	MUS MUSCULUS ME94A12 r1 SOURCE MOUSE EMBRYO 41B433 5 14.5 MUS MUSCULUS cDNA CLONE 40710 5, SIMILAR TO GB H11690
05T2389	gb U05614	2 0e-59	91%	MUS MUSCULUS ME94A12 r1 SOURCE MOUSE EMBRYO 40710 5, SIMILAR TO GB H11690
05T2395	gb AA122609	5 0e-138	10%	MUS MUSCULUS ME94A12 r1 SOURCE MOUSE EMBRYO 40710 5, SIMILAR TO GB H11690
05T2400	gb W14469	4.2e-122	86%	MUS MUSCULUS ME94A12 r1 SOURCE MOUSE EMBRYO 40710 5, SIMILAR TO GB H11690
05T2401	gb AA049859	3 7e-83	91%	MUS MUSCULUS MJ31A01 r1 SOURCE MOUSE EMBRYO 41B433 5 14.5 MUS MUSCULUS cDNA CLONE 475969 5, SIMILAR TO
05T2416	gb W14179	3 1e-54	94%	SW YAG6, 5CHHO Q10106 HYPOPHYSICAL 29 / KD PROTEIN C1366.06 IN CHROMOSOME 1 HOMO SAPIENS A191104 r1 STRATAGENE CLONE 115304, HOMO SAPIENS cDNA CLONE 51203
05T2418	gb H25844	1 3e-41	81%	HOMO SAPIENS YK22102 r1 HOMO SAPIENS cDNA CLONE 267491 5.
05T2431	gb AA104747	2 4e-164	97%	MUS MUSCULUS ME94A12 r1 SOURCE MOUSE EMBRYO 40710 5, SIMILAR TO GB H11690
05T2442	gb W35819	1 7e-55	91%	MUS MUSCULUS ME94A12 r1 SOURCE MOUSE EMBRYO 40710 5, SIMILAR TO GB H11690
05T2447	gb AA061741	1 0e-58	94%	MUS MUSCULUS MJ31A01 r1 SOURCE MOUSE EMBRYO 41B433 5 14.5 MUS MUSCULUS cDNA CLONE 475969 5, SIMILAR TO GB H11690
05T2455	gb AA167801	4 2e-62	90%	HOMO SAPIENS YK22102 r1 HOMO SAPIENS cDNA CLONE 267491 5.
05T2459	gb U05333	3 4e-119	96%	MUS MUSCULUS ME94A12 r1 SOURCE MOUSE EMBRYO 40710 5, SIMILAR TO GB H11690

[illegible]

OST2963	gb w04744	4.2e-31	80%	embryo NMRL13.5 14.5 Mus musculus cDNA clone 479149 5' similar to WP.145112.4 C502740 Homo sapiens zaf9c08 r1 Soares fetal lung NBL19W Homo sapiens cDNA clone 238766 5'
OST2971	gb AA120487	9.2e-107	10%	Mus musculus ml12f07.r1 Beddington mouse embryonic region Mus musculus cDNA clone 537733 5' similar to XE-TIMP2-YEAR2 P38219 HYPOTHETICAL 44.2 KID PROTEIN IN SCOT-MNFI INTERGENIC REGION
OST2974	gb UJ13553	2.6e-102	88%	Nation norvegicus hatus norvegicus Cytosolic protein C precursor mRNA, complete cds
OST2977	gb G97555	6.1e-164	97%	Homo sapiens M musculus MSI Homo sapiens c51b02 r1 Surabagane neuroepithelium [F917231] Homo sapiens cDNA clone 645099 3' similar to TR-G972006 G972006 mRNA; EXPRESSED SEQUENCE TAG
OST2983	gb w49206	1.0e-119	98%	Mus musculus mc91g12 r1 Soares mouse embryo NMRL13.5 14.5 Mus musculus cDNA clone 355942 5' similar to PIR.544900 S44900 ZK652.10 protein - Caenorhabditis elegans
OST2987	gb AA027683	2.3e-134	96%	Mus musculus ml2b01.r1 Soares mouse p3NMFL19.5 Mus musculus cDNA clone 463273 5'
OST2988	gb x52129	2.2e-52	73%	Mus musculus domesticus Mouse testis-specific mRNA pub.62
OST2989	gb AA152050	1.3e-46	78%	Homo sapiens z14Bb12.r1 Soares pregnant uterus NBL19W Homo sapiens cDNA clone 505151 5' similar to gb M90356.cdsl TRANSSCRIPTION FACTOR BPFB3 (HUMAN)
OST2991	gb AA003171	8.4e-151	93%	Mus musculus mg56h09.r1 Soares mouse embryo NMRL13.5 14.5 Mus musculus cDNA clone 437057 5' similar to gb M24194 GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKE PROTEIN (HUMAN);
OST2994	gb X51546	1.9e-51	83%	gb.X75333 N.musculus Homo sapiens y97r12.r1 Homo sapiens cDNA clone 38050 5' similar to SP-VI CLOCCK P02640 Homo sapiens cDNA clone 38050 5' similar to
OST2996	gb x99921	1.6e-82	10%	Mus musculus M musculus mRNA for S100 calcium-binding protein A11
OST2998	gb D19012	3.2e-48	10%	Mus musculus Mouse J.-directed cDNA. MUSGS01209, clone mc0315
OST3003	gb U27502	1.3e-169	97%	Mus musculus Mus musculus lensa major intrinsc protein (MIF) mRNA, complete cds
OST3004	gb AA103385	1.9e-162	98%	Mus musculus mc23f02.r1 Life Tech mouse embryo l3 5dpz 10666014 Mus musculus cDNA clone 554427 5' similar to gb:Z15030_rnal MYOSIN REGULATORY LIGHT CHAIN 2, VENTRICULAR (HUMAN);
OST3011	gb AA035805	1.2e-98	99%	gb:X65979 M.musculus PLMLC-A mRNA for myosin light chain 2 (MOUSE) Mus musculus m15a10.r1 Soares mouse embryo NMRL13.5 11.5 Mus musculus cDNA clone 467226 5' similar to PIR.528237 S28237 NADH dehydrogenase
OST3017	gb AA050908	4.0e-123	92%	Mus musculus mj21o2.r1 Soares mouse embryo NMRL13.5 14.5 Mus musculus cDNA clone 476762 5' similar to SW:AP17.RAT Q00180 CLATHRIN CYT ASSEMBLY PROTEIN AP17
OST3018	gb U083271	2.2e-235	99%	Mus musculus Mons: W/A. of small GTP-binding protein 10, exon2 an' complete cds
OST3032	gb U09305	2.1e-76	99%	Homo sapiens Mus musculus (CP complete cds
OST3035	gb L08651	1.8e-115	90%	Cytochrome b5 reductase (CYPR) mRNA, complete cds
OST3037	gb U09056	4.5e-134	74%	Mus musculus Mus musculus latque ribosomal subunit protein MRNA, complete cds Mus musculus m18b105.r1 Soares mouse embryo NMRL13.5 14.5 Mus musculus cDNA clone 421017 5'

Figure 8 cont'd.

0-TJ305	gb D88433	1 0c-10c	874	Mus musculus House mRNA
0-TJ312	gb D8109	9 1c-5c	664	Mus musculus Mus mRNA
0-TJ323	gb D47643	1 2c-11c	914	prepro-neuturin mRNA, Complete cds
0-TJ324	gb X61399	2 2c-5c	874	Mus musculus House Y1-1 mRNA for Y1-1 protein (nuclear protein with DNA-binding ability), complete cds
0-TJ325	gb D28476	6 5c-10c	944	Mus musculus House Y1-1 mRNA for Y1-1 novel protein
0-TJ349	gb H10210	2 2c-5c	944	gene, complete cds
0-TJ352	gb AA099569	4 9c-6c	774	Mus musculus House Y1-1 mRNA for Y1-1 factor S-11, clone pS11-3
0-TJ354	gb U51638	9 1c-6c	924	pregnant uterus mRNA Homo sapiens cDNA clone AB96731
0-TJ355	gb U49185	4 1c-4c	824	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ366	gb AA122035	2 1c-8c	694	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ370	gb U56105B	4 6c-10c	944	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ371	gb U31107	1 5c-5c	714	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ372	gb U64859	2 2c-13c	994	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ375	gb AA015237	4 0c-4c	104	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ376	gb U21347	4 2c-10c	994	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ388	gb U50264	1 9c-11c	904	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ390	gb U24022	3 6c-16c	784	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ393	gb U60330	1 7c-20c	934	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ404	gb AA168895	6 3c-10c	964	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ413	gb U21837	3 3c-3c	914	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ425	gb U71116	1 3c-10c	884	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ428	gb AA189339	3 4c-37c	884	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ441	gb U51058	7 9c-6c	774	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ450	gb X58426	1 1c-5c	964	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ457	gb U71064	9 0c-16c	974	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ460	gb AA185211	4 2c-13c	994	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ480	gb AA118567	9 4c-10c	894	Mus musculus mRNA Homo sapiens cDNA clone AB96731
0-TJ481	gb X56906	1 0c-17c	954	Mus musculus mRNA Homo sapiens cDNA clone AB96731

OSTJ3483	gb X79446	1 4c-114	924	Mus musculus H. musculus Olf1 mRNA for outer dense fiber protein of sperm tails
OSTJ3485	gb J01024	1.4c-75	864	Human sapiens similar to T cell-specific MA1
OSTJ3492	gb W09518	4.7c-139	924	Mus musculus mab0d09.r1 Soares mouse p1NMf19.5 Mus musculus cDNA clone p30933.5
OSTJ3494	gb W61666	1.1c-138	994	Mus musculus mab2d01.r1 Soares mouse p1NMf19.5 14.5 Mus musculus cDNA clone p74881.5, similar to SW GHG5 NOV19 p10670 GUANTINE NUCLEOTIDE-BINDING PROTEIN
OSTJ3500	gb J062483	2.1c-180	984	G11/G15/G10 GAMMA-5 SUBUNIT, [1] Mus musculus Mus musculus ubiquitin conjugating enzyme (ubc4) mRNA, complete cds
OSTJ3501	gb J59851	6.8c-54	904	Human sapiens human fetal brain cDNA 5'-end GGN-070103
OSTJ3505	gb W40883	3.9c-173	994	Mus musculus mcf39d07.r1 Soares mouse p1NMf19.5 Mus musculus cDNA clone p350893.5
OSTJ3508	gb J023458	2.0c-119	904	Mus musculus Mus musculus endogenous retroviruslike B-26 (distantly related to MuLV) LTR
OSTJ3516	gb J14441	5 4c-177	904	Rattus norvegicus Rat phospholipidylethanolamine N-methyltransferase mRNA, complete cds
OSTJ3517	gb AA015044	5.5c-114	974	Mus musculus mb2f10.r1 Soares mouse placenta 4NMf19.5 14.5 Mus musculus cDNA clone 443371.5
OSTJ3518	gb AA061165	6.3c-99	914	Mus musculus m33f05.r1 Soares mouse embryo NMf19.5 14.5 Mus musculus cDNA clone 477725.5, similar to TR-E222933
OSTJ3521	gb J03756	3.7c-70	874	Rattus norvegicus Rattus norvegicus alpha actinin mRNA, complete cds
OSTJ3531	gb J019893	6.7c-34	804	Bos taurus Bos taurus peptide methionine sulfoxide reductase (mcrA) mRNA, complete cds
OSTJ3534	gb J037150	5.7c-31	834	Mus musculus Mus musculus hemocyt protein (Hox-1.11) gene, complete cds
OSTJ3545	gb J093148	4.0c-103	844	Mus musculus mb48f02.r1 Soares mouse p1NMf19.5 Mus musculus cDNA clone p32667.5
OSTJ3556	gb W08748	1.9c-129	974	Rattus norvegicus Rattus norvegicus (clone RAHD2-5/8) zinc finger protein mRNA, 3' end cds
OSTJ3558	gb J03386	7.9c-132	974	Mus musculus mab94c11.r1 Soares mouse p1NMf19.5 Mus musculus cDNA clone p318356.5, similar to SW-R527-RAT
OSTJ3561	gb W13785	5.1c-64	994	P24051 40S RIBOSOMAL PROTEIN S27, [1] Mus musculus m339d07.r1 Soares mouse embryo NMf19.5 14.5 Mus musculus cDNA clone 478477.5
OSTJ3567	gb AA050004	2.8c-48	784	Mus musculus mcf33a07.r1 Soares mouse embryo NMf19.5 14.5 Mus musculus cDNA clone p31188.5
OSTJ3571	gb W57236	2.4c-113	914	Mus musculus m399a06.r1 Soares mouse p1NMf19.5 Mus musculus cDNA clone p484210.5, similar to gb-X58079.5-100
OSTJ3575	gb AA080212	6.0c-90	934	PROTEIN, ALPHA CHAIN (HUMAN) Gallus gallus PTM 641 subunit-protein phosphatase 1A 21 kDa regulatory subunit, [Rattus norvegicus] Rat 21 kDa smooth muscle, [Mus musculus] Mus musculus, [Rattus norvegicus] Rattus norvegicus XPAC Xaphima protein-coding group A Collecting gene, accession 6
OSTJ3582	gb X74350	1 5c-74	994	Human sapiens Human creatine nucleotide exchange factor p5422 mRNA, complete cds
OSTJ3601	gb J050018	4 6c-138	894	Human sapiens Y48H05.r1 Homo sapiens cDNA clone Z9481.5
OSTJ3602	gb J015062	2 4c-107	904	Rattus norvegicus Rat 24-kDa subunit of mitochondrial NADH dehydrogenase, [Homo sapiens] Homo sapiens
OSTJ3604	gb J022756	4 9c-119	844	mRNA, 3' end
OSTJ3608	gb J011994	5 4c-101	854	Human sapiens Human DAP-dependent

[illegible]

OSI1188	gb AA014426	9 7e-55	101	Mus musculus mbl4001.11 Soares mouse embryo NMRE13.5 14.5 Mus musculus cDNA 54 NM7M NOV1M 021267 MAMU-UNIQTHONE OXIDOLACTONASE-B17 SUBUNIT
OST3189	gb D13544	9.5e-67	971	Mus musculus Mouse cDNA for primase small subunit complete cds
OST3807	gb W26968	3.8e-51	801	Homo sapiens 16k7 Human retina cDNA randomly primed sublibrary Homo sapiens cDNA
OST3818	gb W20248	3.8e-48	961	unidentified cloning vector Moloney murine leukemia virus retroviral vector pLXSN, complete genome
OST3819	gb T556J2	3 0e-35	811	Human placenta cDNA 11 Homo sapiens cDNA clone 71517.5 similar to SP T2065.10 C2A0629
OST3827	gb AA046830	1 2e-67	841	Homo sapiens c12h11.11 Soares fetal heart MAM19W Homo sapiens cDNA clone 37673.1
OST3831	gb W70777	3.5e-121	991	Mus musculus m4440.2.11 Soares mouse embryo NMRE13.5 14.5 Mus musculus cDNA
OST3839	gb W86008	1 4e-103	861	Homo sapiens EST025313 Homo sapiens cDNA clone HPKCY19 similar to hypothetical 43.1K protein
OST3843	gb Z82190	2 8e-51	881	HOMO SAPIENS HUMAN DNA SEQUENCE *** SEQUENCING IN PROGRESS ** from clone 110M12. HGS phase
OST3849	gb W64986	1 3e-173	941	Mus musculus m40405.11 Soares mouse embryo NMRE13.5 14.5 Mus musculus cDNA clone 386504.5 similar to SW-V3H7 DIC01 P14327 VIGETATVIL
OST3851	gb U51037	1 0e-135	841	SPLICIFIC PROPLIN H7 [1] Mus musculus Mus transcription factor (CTCF) mRNA, complete cds
OST3858	gb X56135	4 7e-237	971	Mus musculus Mouse mRNA for prothymosin alpha
OST3864	gb D119493	9 8e-13	951	Mus musculus Mouse J-cyretic cDNA, RUGS008B1, clone m40610
OST3869	gb W41525	4 4e-100	851	Mus musculus m45104.11 Soares mouse p1NMF19.5 Mus musculus cDNA clone 351439.5
OST3897	gb W10485	3.8e-97	951	Mus musculus m45306.11 Soares mouse p1NMF19.5 Mus musculus cDNA clone 314434.5
OST3903	gb W93988	1 2e-108	801	Mus musculus m47102.11 Soares mouse embryo NMRE13.5 14.5 Mus musculus cDNA clone 374619.5 similar to gb U07151
OST3905	gb W05430	8 0e-102	921	ADP-RIBOSYLATION FACTOR-LIKE PROTEIN 3 (HUMAN)
OST3909	gb AA020459	1 2e-80	941	Mus musculus Mouse Muf1 mRNA, exon 5
OST3917	gb Z44044	0.7e-81	871	Mus musculus m46106.11 Soares mouse placenta 4HMF13.5 14.5 Mus musculus cDNA clone 455410.5
OST3924	gb J004699	3.9e-32	841	Homo sapiens H. sapiens partial cDNA sequence, clone c-18c07
OST3925	gb W23511	1 2e-88	761	Mus musculus Mouse nicotinic acetylcholine receptor beta subunit (nAChR2) gene, complete cds
OST3931	gb U14957	1.6e-36	811	Homo sapiens Z46602.11 Soares fetal lung MAM19W Homo sapiens cDNA clone 306630.5
OST3945	gb W15004	1 6e-122	971	Homo sapiens Human 53K isoform of Type 1 phosphatidylinositol-4-phosphate 5-kinase (PIP5K) mRNA, complete cds
OST3957	gb AA051293	2 8e-143	961	Mus musculus m45209.11 Soares mouse embryo NMRE13.5 14.5 Mus musculus cDNA clone 478627.5 similar to SW T2065.10 C2A0629
OST3960	gb Z88614	1 1e-88	821	GROWTH FACTOR BETA-1 BINDING PROTEIN PRICUSON
OST3961	gb UC7988	6 6e-37	771	Mus musculus Mouse 921-5 mRNA for presynaptic protein, complete cds
				Homo sapiens Human synaptobrevin-1 associated protein (SNAP) mRNA, complete cds

Figure 8 cont'd.

OST4196	gb w41301	3 1e-39	99%	Mus musculus mcd1006.r1 Soares mouse p18f12.5
OST4201	gb AA203787	2.7e-09	90%	Mus musculus muf0612.r1 Soares mouse clone 643873.5
OST4220	gb 551016	9.3e-205	92%	Mus musculus NM1613.5
OST4229	gb Z31263	4.8e-70	97%	bov tauris E2125K2=multibiquitinating enzyme [cattle, thymus, mRNA, 825 nt]
OST4235	gb w53187	3.0e-173	97%	Mus musculus M.musculus expressed sequence tag M7857
OST4243	gb AA048921	2.3e-40	86%	Mus musculus md19a07.r1 Soares mouse embryo NM1613.5 14.5 Mus musculus cdna clone 368820.5' similar to MF.C32D5.9 CE01849
OST4243	gb AA048921	2.3e-40	86%	Mus musculus m374ell.r1 Soares mouse embryo NM1613.5 14.5 Mus musculus cdna clone 479276.5' similar to gb:013705
OST4245	gb 010216	9.9e-80	75%	Mus musculus domesticus C57BL/6J plama glutathione (H0D5E)
OST4247	gb AA023146	1.5e-115	96%	Homo sapiens nm02035.s1 Homo sapiens CDNA clone 47679.5'
OST4251	gb AA070774	8.7e-154	98%	Mus musculus m67d03.r1 Soares mouse placenta 48Mw11.5 14.5 Mus musculus CDNA clone 455981.5' similar to SW:AAIP.HUMAN.Q04941.INTESTINAL.MEMBRANE.A4.PROTEIN.[1]
OST4254	gb w54737	2.4e-02	10%	Homo sapiens nm53911.s1 Scratogene fibroblast (#937212) Homo sapiens cdna clone 529412.3'
OST4258	gb AA013789	4.3e-169	90%	Mus musculus md10a04.r1 Soares mouse embryo NM1613.5 14.5 Mus musculus cdna clone 367950.5'
OST4281	gb 016175	4.0e-40	63%	Mus musculus mhl3d03.r1 Soares mouse placenta 48Mw11.5 14.5 Mus musculus cdna clone 42473.5' similar to PIR:JC2477.4.147120.00 Protein - human fibroblast clone 3 (Vhs33) gene, partial fibroblastin 3 (Vhs33) gene, complete cds
OST4283	gb AA007519	8.9e-52	81%	Homo sapiens zn19812.r1 Soares fetal liver spleen INFL5 S1 Homo sapiens cdna clone 429358.5'
OST4288	gb AA000024	1.4e-115	96%	Mus musculus mg3j406.r1 Soares mouse embryo NM1613.5 14.5 Mus musculus cdna clone 425602.5' similar to gb:XD3920.fma2 M.musculus GSH1X gene (H0D261)
OST4315	gb 010210	6.4e-62	96%	Mus musculus Mouse transcripition factor 5-11, clone p511-3
OST4319	gb 304696	2.0e-127	95%	Mus musculus Mouse p511-3 S-transferrase classa mu (GST5-5) mRNA, complete cds

05T41971	gb W45926	9.6e-55	941	Mus musculus mc79c04.r1 Soares mouse embryo NM0E11.5 14.5 Mus musculus cDNA clone 354750 5'.
	gb W13524	2.6e-111	901	pseudogene (psi-SAA)
	gb W16778	4.7e-45	831	homo sapiens YF31A08.s1 Homo sapiens cDNA clone 128630 3'.
	gb AA000314	1.9e-112	961	Mus musculus mg44c07.r1 Soares mouse embryo NM0E11.5 14.5 Mus musculus cDNA clone 425700 5'.
	gb U37297	2.9e-121	911	Mus musculus Mus musculus (clone B6) myeloid secondary granule protein mRNA
	gb U26664	2.0e-155	941	Mus musculus EST F032
	gb U007470	7.5e-93	921	Sequence tag EST F032
	gb AA084704	2.2e-54	881	gene partial cds
	gb U354033	2.2e-54	881	homo sapiens zn05f04.s1 Stralagene hntf neuron (1937233) homo sapiens cDNA clone 546559 3' similar to TR:G600529 G600529 NADH UBIQUINONE OXIDOREDUCTASE SUBUNIT
	gb F03500	7.6e-63	861	homo sapiens H. sapiens partial cDNA sequence; clone c-12d08
	gb WJ0618	3.1e-118	971	Mus musculus mc10h12.r1 Soares mouse p1NM19.5 Mus musculus cDNA clone 348167 5'.
	gb WJ6515	6.0e-135	941	Mus musculus mb76g12.r1 Soares mouse p1NM19.5 Mus musculus cDNA clone 335398 5'.
	gb XB2021	2.0e-105	911	Rattus norvegicus R.norvegicus mRNA for heat shock related protein
	gb U63704	3.3e-140	861	Rattus norvegicus complete cds
	gb W75804	1.1e-84	931	Mus musculus mc67406.r1 Soares mouse embryo NM0E11.5 14.5 Mus musculus cDNA clone 400504 5'.
	gb W20730	6.5e-90	961	Mus musculus mb56g01.r1 Soares mouse p1NM19.5 Mus musculus cDNA clone 337296 5'.
	gb AA042474	2.4e-33	691	homo sapiens zk34h10.s1 Soares pregnant uterus NB1P0 Homo sapiens cDNA clone 486677 3'.
	gb U31489	3.0e-84	851	Rattus sp. EST105564 Rattus sp. cDNA 3' end
	gb W71052	3.7e-121	911	Mus musculus me27f01.r1 Soares mouse embryo NM0E11.5 14.5 Mus musculus cDNA clone 400504 5'.
				Cloned from a rat testis cDNA library using yeast 3' RARE HYPOPHYSICAL 13.6 kb PROTEIN IN PBT112-ILS1 IMPROGENIC REGION. 11
	gb C07091	5.7e-74	891	Rattus norvegicus similar to none
	gb X56135	4.4e-41	811	Mus musculus Mouse mRNA for prothymosin alpha
	gb W54510	1.5e-135	911	embryo NM0E11.5 14.5 Mus musculus cDNA clone 367841 5' similar to PIR:A56059 A56059 protein-tyrosine-phosphatase
	gb U36393	2.6e-111	961	Mus musculus Mus musculus transcription factor TFEb mRNA, partial cds
	gb X56046	1.7e-161	961	Mus musculus House mRNA (clone lambda-16) for hypothetical protein A
	gb X05900	3.5e-58	851	Rattus norvegicus Rattus norvegicus beta1-crystallin (beta1-cryst)
	gb U53859	8.0e-169	901	Rattus norvegicus Rattus norvegicus beta1 subunit (beta1) mRNA, partial cds
	gb U61195	1.3e-10	841	Mus musculus Mus musculus x inactive specific transcript (Xlat) gene, clemid M4-14A, fragment 2
	gb X63507	2.0e-75	811	Mus musculus H.musculus H0X-3.5 gene
	gb W85357	2.2e-83	821	Mus musculus mf9h12.r1 Soares mouse embryo NM0E11.5 14.5 Mus musculus cDNA clone 408455 5' similar to SW GLYH_HUMAN P34897 SERINE HYDROXYMETHYLTRANSFERASE, MITOCHONDRIAL
	gb W11463	8.9e-38	871	Mus musculus mc11e07.r1 Soares mouse p1NM19.5 Mus musculus cDNA clone 340148 5'.

DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below at 201 et seq. underneath my name.

I believe I am the original, first and sole inventor if only one name is listed at 201 below, or an original, first and joint inventor if plural names are listed at 201 et seq. below, of the subject matter which is claimed and for which a patent is sought on the invention entitled

AN INDEXED LIBRARY OF CELLS CONTAINING GENOMIC MODIFICATIONS AND METHODS OF MAKING AND UTILIZING THE SAME

and for which a patent application:

- ☐ is attached hereto and includes amendment(s) filed on _____ (if applicable)
- ☒ was filed in the United States on October 2, 1997 as Application No. _____ (for declaration not accompanying application)
- with amendment(s) filed on _____ (if applicable)
- ☐ was filed as PCT international Application No. _____ on _____ and was amended under PCT Article 19 on _____ (if applicable)

I hereby state that I have reviewed and understand the contents of the above identified application, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119(a)-(d) of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

EARLIEST FOREIGN APPLICATION(S), IF ANY, FILED PRIOR TO THE FILING DATE OF THE APPLICATION			
APPLICATION NUMBER	COUNTRY	DATE OF FILING (day, month, year)	PRIORITY CLAIMED
			YES <input type="checkbox"/> NO <input type="checkbox"/>
			YES <input type="checkbox"/> NO <input type="checkbox"/>

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

APPLICATION NUMBER	FILING DATE

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code §112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

APPLICATION SERIAL NO.	FILING DATE	STATUS		
		PATENTED	PENDING	ABANDONED
08/726,867	October 4, 1996		X	
08/728,963	October 11, 1996		X	
08/907,598	August 8, 1997		X	

POWER OF ATTORNEY: As a named inventor, I hereby appoint S. Leslie Misrock (Reg. No. 18872), Harry C. Jones, III (Reg. No. 20280), Berj A. Terzian (Reg. No. 20060), Gerald J. Flintoft (Reg. No. 20823), David Weild, III (Reg. No. 21094), Jonathan A. Marshall (Reg. No. 24614), Barry D. Rein (Reg. No. 22411), Stanton T. Lawrence, III (Reg. No. 25736), Isaac Jarkovsky (Reg. No. 22713), Joseph V. Colaianni (Reg. No. 20019), Charles E. McKenney (Reg. No. 22795), Philip T. Shannon (Reg. No. 24278), Francis E. Morris (Reg. No. 24615), Charles E. Miller (Reg. No. 24576), Gidon D. Stern (Reg. No. 27469), John J. Lauter, Jr. (Reg. No. 27814), Brian M. Poissant (Reg. No. 28462), Brian D. Coggio (Reg. No. 27624), Rory J. Radding (Reg. No. 28749), Stephen J. Harbulak (Reg. No. 29166), Donald J. Goodell (Reg. No. 19766), James N. Palik (Reg. No. 25510), Thomas E. Friebe (Reg. No. 29258), Laura A. Coruzzi (Reg. No. 30742), Jennifer Gordon (Reg. No. 30753), Jon R. Stark (Reg. No. 30111), Allan A. Fanucci (Reg. No. 30256), Geraldine F. Baldwin (Reg. No. 31232), Victor N. Balancia (Reg. No. 31231), Samuel B. Abrams (Reg. No. 30605), Steven I. Wallach (Reg. No. 35402), Marcia H. Sundeen (Reg. No. 30893), Paul J. Zegger (Reg. No. 33821), Edmond R. Bannon (Reg. No. 32110), Bruce J. Barker (Reg. No. 33291), Adriane M. Antler (Reg. No. 32605), Ann L. Gisolfi (Reg. No. 31956), Mark A. Farley (Reg. No. 33170), and James G. Markey (Reg. No. 31636), all of Pennie & Edmonds LLP, whose addresses are 1155

Avenue of the Americas, New York, New York 10036, 1667 K Street N.W., Washington, DC 20006 and 3300 Hillview Avenue, Palo Alto, CA 94304, and each of them, my attorneys, to prosecute this application, and to transact all business in the Patent and Trademark Office connected therewith.

SEND CORRESPONDENCE TO: PENNIE & EDMONDS LLP 1155 AVENUE OF THE AMERICAS NEW YORK, N.Y. 10036-2711				DIRECT TELEPHONE CALLS TO: PENNIE & EDMONDS LLP DOCKETING (212) 790-2803	
201	FULL NAME OF INVENTOR	LAST NAME Zambrowicz	FIRST NAME Brian	MIDDLE NAME	
	RESIDENCE & CITIZENSHIP	CITY The Woodlands	STATE OR FOREIGN COUNTRY TX	COUNTRY OF CITIZENSHIP USA	
	POST OFFICE ADDRESS	STREET 18 Firethorne Place	CITY The Woodlands	STATE OR COUNTRY TX	ZIP CODE 77382
202	FULL NAME OF INVENTOR	LAST NAME Friedrich	FIRST NAME Gienn	MIDDLE NAME A.	
	RESIDENCE & CITIZENSHIP	CITY The Woodlands	STATE OR FOREIGN COUNTRY TX	COUNTRY OF CITIZENSHIP Canada	
	POST OFFICE ADDRESS	STREET 30 Reflection Point	CITY The Woodlands	STATE OR COUNTRY TX	ZIP CODE 77381
203	FULL NAME OF INVENTOR	LAST NAME Bradley	FIRST NAME Allan	MIDDLE NAME	
	RESIDENCE & CITIZENSHIP	CITY Houston	STATE OR FOREIGN COUNTRY TX	COUNTRY OF CITIZENSHIP England	
	POST OFFICE ADDRESS	STREET 5127 Queensloch	CITY Houston	STATE OR COUNTRY TX	ZIP CODE 77096
204	FULL NAME OF INVENTOR	LAST NAME Sands	FIRST NAME Arthur	MIDDLE NAME T.	
	RESIDENCE & CITIZENSHIP	CITY The Woodlands	STATE OR FOREIGN COUNTRY TX	COUNTRY OF CITIZENSHIP USA	
	POST OFFICE ADDRESS	STREET 163 Bristol Bend Circle	CITY The Woodlands	STATE OR COUNTRY TX	ZIP CODE 77382
205	FULL NAME OF INVENTOR	LAST NAME	FIRST NAME	MIDDLE NAME	
	RESIDENCE & CITIZENSHIP	CITY	STATE OR FOREIGN COUNTRY	COUNTRY OF CITIZENSHIP	
	POST OFFICE ADDRESS	STREET	CITY	STATE OR COUNTRY	ZIP CODE
206	FULL NAME OF INVENTOR	LAST NAME	FIRST NAME	MIDDLE NAME	
	RESIDENCE & CITIZENSHIP	CITY	STATE OR FOREIGN COUNTRY	COUNTRY OF CITIZENSHIP	
	POST OFFICE ADDRESS	STREET	CITY	STATE OR COUNTRY	ZIP CODE

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

SIGNATURE OF INVENTOR 201 <i>Brian Zambrowicz</i> DATE 10/16/97	SIGNATURE OF INVENTOR 202 <i>Gienn A. Friedrich</i> DATE 10.16.97	SIGNATURE OF INVENTOR 203 <i>Allan Bradley</i> DATE 2-17-98
SIGNATURE OF INVENTOR 204 <i>Arthur T. Sands</i> DATE 10/16/97	SIGNATURE OF INVENTOR 205 <i>[Signature]</i> DATE 2-17-98	SIGNATURE OF INVENTOR 206 <i>[Signature]</i> DATE

PATENT

Attorney Docket No. 07705.0002-01000

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of)

Glenn A. Friedrich *et al.*)

Serial No.: 08/942,806)

) Group Art Unit: 1815

Filed: October 2, 1997)

) Examiner: Unknown

For: AN INDEXED LIBRARY OF CELLS)
CONTAINING GENOMIC MODIFICATIONS)
AND METHODS OF MAKING AND)
UTILIZING THE SAME)

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

**REVOCATION OF POWER OF ATTORNEY
AND GRANT OF NEW POWER OF ATTORNEY**

The undersigned, a representative authorized to sign on behalf of the assignee owning all of the interest in this patent application, hereby revokes all previous powers of attorney or authorization of agent granted in this application before the date of execution hereof. The undersigned verifies that Lexicon Genetics, Inc., 4000 Research Forest Drive, The Woodlands, Texas 77381, is the assignee of the entire right, title, and interest in the patent application identified above by virtue of an assignment from the inventors recorded in the U.S. Patent and Trademark Office at Reel 9215, Frame 0835.

The undersigned certifies that the evidentiary documents have been reviewed and to

LAW OFFICES
NNECAN, HENDERSON,
FARABOW, GARRETT
& DENNER, L.L.P.
MANFORD RESEARCH PARK
700 HANSEN WAY
SALO ALTO, CALIF. 94501
920-840-5800

the best of the undersigned's knowledge and belief, title is in the assignee Lexicon Genetics, Inc.

The undersigned hereby grants its power of attorney to **FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, L.L.P.**, Douglas B. Henderson, Reg. No. 20,291; Ford F. Farabow, Jr., Reg. No. 20,630; Arthur S. Garrett, Reg. No. 20,338; Donald R. Dunner, Reg. No. 19,073; Brian G. Brunsvold, Reg. No. 22,593; Tipton D. Jennings, IV, Reg. No. 20,645; Jerry D. Voight, Reg. No. 23,020; Laurence R. Hefter, Reg. No. 20,827; Kenneth E. Payne, Reg. No. 23,098; Herbert H. Mintz, Reg. No. 26,691; C. Larry O'Rourke, Reg. No. 26,014; Albert J. Santorelli, Reg. No. 22,610; Michael C. Elmer, Reg. No. 25,857; Richard H. Smith, Reg. No. 20,609; Stephen L. Peterson, Reg. No. 26,325; John M. Romary, Reg. No. 26,331; Bruce C. Zotter, Reg. No. 27,680; Dennis P. O'Reilly, Reg. No. 27,932; Allen M. Sokal, Reg. No. 26,695; Robert D. Bajefsky, Reg. No. 25,387; Richard L. Stroup, Reg. No. 28,478; David W. Hill, Reg. No. 28,220; Thomas L. Irving, Reg. No. 28,619; Charles E. Lipsey, Reg. No. 28,165; Thomas W. Winland, Reg. No. 27,605; Basil J. Lewis, Reg. No. 28,818; Martin I. Fuchs, Reg. No. 28,508; E. Robert Yoches, Reg. No. 30,120; Barry W. Graham, Reg. No. 29,924; Susan Haberman Griffen, Reg. No. 30,907; Richard B. Racine, Reg. No. 30,415; Thomas H. Jenkins, Reg. No. 30,857; Robert E. Converse, Jr., Reg. No. 27,432; Clair X. Mullen, Jr., Reg. No. 20,348; Christopher P. Foley, Reg. No. 31,354; Roger D. Taylor, Reg. No. 28,992; John C. Paul, Reg. No. 30,413; David M. Kelly, Reg. No. 30,953; Kenneth J. Meyers, Reg. No. 25,146; Carol P. Einaudi, Reg. No. 32,220; Walter Y. Boyd, Jr., Reg. No. 31,738; Steven M. Anzalone, Reg. No. 32,095; Jean B. Fordis, Reg. No. 32,984; Barbara C. McCurdy, Reg.

No. 32,120; James K. Hammond, Reg. No. 31,964; Richard V. Burgujian, Reg. No. 31,744; J. Michael Jakes, Reg. No. 32,824; Thomas W. Banks, Reg. No. 32,719; Christopher P. Isaac, Reg. No. 32,616; Bryan C. Diner, Reg. No. 32,409; M. Paul Barker, Reg. No. 32,013; Andrew Chanhon Sonu, Reg. No. 33,457; David S. Forman, Reg. No. 33,694; Vincent P. Kovalick, Reg. No. 32,867; James W. Edmondson, Reg. No. 33,871; Michael R. McGurk, Reg. No. 32,045; Joann M. Neth, Reg. No. 36,363; Gerson S. Panitch, Reg. No. 33,751; Cheri M. Taylor, Reg. No. 33,216; Charles E. Van Horn, Reg. No. 40,266; Linda A. Wadler, Reg. No. 33,218; Jeffrey A. Berkowitz, Reg. No. 36,743; Michael R. Kelly, Reg. No. 33, 921; James B. Monroe, Reg. No. 33,971; and Norbert Stahl, Reg. No. 44,350; both jointly and separately as their attorneys with full power of substitution and revocation to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith, and to receive the Letters Patent.

Please send all future correspondence concerning this application to Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P. at the following address:

Finnegan, Henderson, Farabow,
Garrett & Dunner, L.L.P.
1300 I Street, N.W.
Washington, D.C. 20005-3315

Dated: 11/22/99

By: 

Lance Ishimoto 41866
Vice President of Intellectual Property
Lexicon Genetics, Inc.